OM of: US-09-301-906-3 to: A_Geneseq_36:* out_format : pfs

Date: Jan 31, 2001 3:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09301906/runat_31012001_153034_9577/app_query.fasta_1.6881
-DB=A_Geneseq_36 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09301906_@CGN1_1_396 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-301-906-3
Query length: 6714
Database: A_Geneseq_36:*
Database sequences: 268485
Database length: 34193795
Search time (sec): 277.730000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58148	+	11520.00	19619.90	0.0	2237	!
/SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:W21636	+	5247.00	8908.82	0.0	1390	! G
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58145	+	1495.00	2520.59	3.1e-133	291	! GL
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58144	+	1432.00	2413.92	2.9e-127	272	! GL
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W73476	+	802.00	1305.05	1.8e-66	2639	! G
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y91969	+	631.50	1044.38	4.9e-51	309	! PM
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58143	+	605.00	1012.62	7.5e-49	120	! GL
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y91963	+	323.00	507.85	2.0e-21	598	! PM
/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R05107	+	323.00	498.79	3.3e-21	1128	! S
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R60608	+	309.00	475.02	7.1e-20	1116	! T
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W87724	+	290.00	433.12	7.9e-18	2161	! R
/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R06531	+	274.50	417.76	1.2e-16	992	! Pr
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R49657	+	237.50	346.81	6.4e-13	1704	! S
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y31381	+	186.50	259.72	4.6e-08	1698	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y31384	+	185.00	257.08	6.4e-08	1708	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W56322	+	185.00	254.55	7.4e-08	2039	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W76368	+	184.50	256.35	7.1e-08	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W93408	+	181.00	250.24	1.5e-07	1708	! S
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R51264	+	179.50	247.81	2.1e-07	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:R91813	+	179.50	247.81	2.1e-07	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W81519	+	179.50	247.81	2.1e-07	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R60562	-	179.00	248.33	2.2e-07	1537	! Y
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W80196	+	169.50	230.72	1.9e-06	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W71209	+	169.50	230.72	1.9e-06	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:W03740	-	160.50	217.25	1.2e-05	1481	! I
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W95160	-	160.50	217.25	1.2e-05	1481	! A
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W89299	-	160.50	217.25	1.2e-05	1481	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y78508	-	160.50	217.25	1.2e-05	1481	! A
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y68447	-	160.50	217.25	1.2e-05	1481	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:R14618	+	160.50	215.34	1.4e-05	1693	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:W06725	-	158.50	221.79	1.2e-05	849	! FL
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W47273	+	158.00	209.08	2.6e-05	1946	! L
/SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:P93357	+	154.50	206.61	4.6e-05	1522	! S
/SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:P94365	+	152.50	201.57	7.9e-05	1705	! S
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y59288	-	151.50	208.12	6.1e-05	957	! Hu
/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R08031	+	151.50	199.86	9.8e-05	1706	! A
/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R05041	+	150.00	186.44	0.0003	3647	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R47578	-	149.50	205.67	8.9e-05	894	! Fl
/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R58754	-	149.50	205.67	8.9e-05	894	! S.
/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y87407	+	147.00	181.51	0.0005	3596	! H
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W88417	+	146.50	200.01	0.0002	928	!
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y35060	+	146.50	199.69	0.0002	949	!
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y37825	+	146.50	196.41	0.0002	1194	
/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y05477	-	145.00	200.49	0.0002	750	!
/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:W41935	+	145.00	193.39	0.0003	1233	

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58148

seq_documentation_block:

```
ID   Y58148 standard; Protein; 2237 AA.
XX
AC   Y58148;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
FH   Key             Location/Qualifiers
FT   Domain          85..204
FT                   /note= "Proteinase domain (Y58143)"
FT   Domain          460..731
FT                   /note= "Methyltransferase domain (Y58144)"
FT   Domain          1922..2212
FT                   /note= "Helicase domain (Y58145)"
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49208.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 3; Fig 16; 84pp; English.
XX
CC   This sequence represents the polyprotein encoded by open reading
CC   frame (ORF) 1a from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 1l.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
```

```
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   2237 AA;


alignment_scores:
           Quality: 11520.00              Length:   2237
             Ratio:  5.150                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-301-906-3 x Y58148   ..

 Align seg 1/1  to: Y58148  from: 1  to: 2237

       1 ATGGACTACATTCGCCCATTGCGCGTTTTCTCCTTTCCTCACGTTAATAA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetAspTyrIleArgProLeuArgValPheSerPheProHisValAsnAs 17

      51 CACCTTGGAGTACGTTAGGTACAACAAGGCCAATGGTGATGTAGGAGCTT 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 nThrLeuGluTyrValArgTyrAsnLysAlaAsnGlyAspValGlyAlaP 34

     101 TCCTAACCACCATGAAGTTCATAGGGAACGTGAAGTTGTCGGACTTCACA 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 heLeuThrThrMetLysPheIleGlyAsnValLysLeuSerAspPheThr 50

     151 CCCAGGTGCGCAGCTATGATTTACATTGGAAAGCTCACCAAAGGGGTGAA 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ProArgCysAlaAlaMetIleTyrIleGlyLysLeuThrLysGlyValLy 67

     201 GCGTACGTTTGTCCCCCCACCAGTTAAAGGGTTTGCACGGCAGTACGCTG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 sArgThrPheValProProProValLysGlyPheAlaArgGlnTyrAlaV 84

     251 TTGTCAGCGGCTCAGTCAGCGCGCTGAGAGGGGATGGTAAGAAGGTCTTG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 alValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeu 100

     301 ATGGAGGCAAGGACCTCAACTTCCGCAACTTCCGACGTGTCTGATTTCGA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 MetGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAs 117

     351 CGTCGTATTCGAAGCTGTTTCTAATGCATTACTTGTCGTACACTACCACC 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 pValValPheGluAlaValSerAsnAlaLeuLeuValValHisTyrHisA 134

     401 GGGTAGTGCCGTATGCCCCCGTCAAGCGCGAGCAGCCTAAACCGGCTGTT 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgValValProTyrAlaProValLysArgGluGlnProLysProAlaVal 150

     451 AAGCAAGATGAGCAGAAGCCCAAACGGCAAGCGTCACATTGGGCTGTTAA 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 LysGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLy 167

     501 GCCAACAGCTGTTGGCGTCCACGTACCACTTCCTAAAAAACAGGAAGCAC 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 sProThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaL 184

     551 TGGAGCCAGCGCAATCAGTCCCACAACAGTCGTTGGAGGAGAAGGCCGCC 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euGluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAla 200

     601 TTGACGTTTGGCCTTTTCTTCAGTAAAGGTGGGGGTGATGAGAGCGACGC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 LeuThrPheGlyLeuPhePheSerLysGlyGlyGlyAspGluSerAspAl 217

     651 TGTCATCTTGCGGAAAGGGAAATTGTTTAACAGGGCCCTTAATGTTCCTA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 aValIleLeuArgLysGlyLysLeuPheAsnArgAlaLeuAsnValProI 234

     701 TTGATGTAAAGAACACGTTCGTTTGGGCTAAAATCTGGGATGAAGCCTCT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 leAspValLysAsnThrPheValTrpAlaLysIleTrpAspGluAlaSer 250

     751 CGTAGGAGAGGGTATTTTTACGTCAAAGATAGAGCTGTTAAATTCTTCCC 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ArgArgArgGlyTyrPheTyrValLysAspArgAlaValLysPhePhePr 267

     801 TATTGTGCGGGGTAGGGCTACGATCGAGGACTTCATCGTGAATACAGCCC 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 oIleValArgGlyArgAlaThrIleGluAspPheIleValAsnThrAlaP 284

     851 CAGGGTGTGATGTTGCCTTGCCGCGCATTGAGTTGTGGAGTATGCGCGAA 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 roGlyCysAspValAlaLeuProArgIleGluLeuTrpSerMetArgGlu 300

     901 AGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAACAATGAGAG 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 ArgAlaPheValCysThrThrLysGlyTrpCysTrpPheAsnAsnGluAr 317

     951 GCTGAGGGGAGAAATTTACAGACGTCGTTGCTTCTCATCTTCCTTTTCGA 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 gLeuArgGlyGluIleTyrArgArgArgCysPheSerSerSerPheSerI 334

    1001 TAGGTTTCTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTT 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 leGlyPheLeuMetHisLeuGlyPheArgSerLeuLysValIleArgPhe 350

    1051 GCGGGCACGAACATACTACACATGCCATCACTCAATGAAGAGCGTACCTT 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 AlaGlyThrAsnIleLeuHisMetProSerLeuAsnGluGluArgThrPh 367

    1101 TGGGTGGAAGGGCGGAGACGTCTATCTCCCCAATGTCCCAAAAACCGCTA 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 eGlyTrpLysGlyGlyAspValTyrLeuProAsnValProLysThrAlaI 384

    1151 TCGTCGCTGGCGATAGGACACGGTTGGGAGGGGAGATCTTGGCCTCCGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 leValAlaGlyAspArgThrArgLeuGlyGlyGluIleLeuAlaSerVal 400

    1201 GCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTAT 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 AlaAsnAlaLeuAsnGlnGluGluValTyrSerSerValValSerSerIl 417

    1251 CACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGG 1300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 eThrAsnArgLeuValLeuArgAspGlnSerAlaLeuLeuSerHisLeuA 434

    1301 ACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAA 1350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     434 spThrLysLeuCysAspMetPheSerGlnArgAspAlaMetIleArgGlu 450

    1351 AAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAA 1400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 LysProSerHisArgCysAspValPheLeuLysProArgGluArgGluLy 467

    1401 GCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCA 1450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     467 sLeuArgGluLeuPheProGluLeuSerIleGlnPheSerAspSerValA 484

    1451 GGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     484 rgSerSerHisProPheAlaAsnAlaMetArgSerCysPheAsnGlyIle 500

    1501 TTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTT 1550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 PheSerArgArgCysGlyAsnValCysPhePheAspIleGlyGlySerPh 517

    1551 CACGTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAG 1600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     517 eThrTyrHisValLysAlaGlyHisValAsnCysHisValCysAsnProV 534
```

```
1601 TCCTAGACGTTAAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTT 1650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 alLeuAspValLysAspValLysArgArgIleAsnGluIleLeuPheLeu 550

1651 TCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGC 1700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 SerThrAlaGlyGlyAspSerTyrValSerSerAspLeuLeuThrGluAl 567

1701 GGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATT 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 aAlaSerLysSerValSerTyrCysSerArgGluSerGlnAsnCysAspS 584

1751 CTAGAGCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAG 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 584 erArgAlaAspAlaGlyPheMetValAspValTyrAspIleSerProGln 600

1801 CAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGC 1850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 GlnValAlaGluAlaMetAspLysLysGlyAlaLeuValPheAspIleAl 617

1851 TCTTATGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGG 1900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 617 aLeuMetPheProValGluLeuLeuTyrGlyAsnGlyGluValTyrLeuG 634

1901 AAGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAAT 1950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 634 luGluLeuAspThrLeuValLysArgGluGlyAspTyrLeuAlaTyrAsn 650

1951 GTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGG 2000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 651 ValGlyGlnCysGlyGluMetTyrGluHisSerPheSerAsnValSerGl 667

2001 GTTTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGC 2050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 667 yPhePheThrPheSerTyrValArgThrSerSerGlyAsnValPheLysL 684

2051 TAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGG 2100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 684 euGluTyrGluGlyTyrArgCysGlyTyrHisHisLeuThrMetCysArg 700

2101 GCTCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTC 2150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 701 AlaGlnLysSerProGlyThrGluValThrTyrArgSerLeuValProSe 717

2151 GTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTG 2200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 717 rPheValGlyLysSerLeuValPheIleProValValAlaGlySerServ 734

2201 TGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTAT 2250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 734 alSerPheLysThrIleValLeuAspSerAspPheValAspArgIleTyr 750

2251 TCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGTA 2300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 SerTyrAlaLeuAsnThrIleGlyThrPheGluAsnArgThrPheGluTy 767

2301 TGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGTC 2350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 767 rAlaValGlyAlaValArgSerGlnLysThrHisValIleThrGlySerA 784

2351 GCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTA 2400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 784 rgValValHisSerLysValAspIleSerProAspAspMetTrpGlyLeu 800

2401 GTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTATTCG 2450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 ValValAlaValMetAlaGlnAlaIleLysAspArgAlaLysSerIleAr 817

2451 CTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCA 2500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 817 gSerTyrAsnPheIleLysAlaSerGluGlySerLeuAlaGlyValPheL 834

2501 AGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTC 2550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 834 ysLeuPhePheGlnThrValGlyAspCysPheSerAsnAlaValSerVal 850

2551 TATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGCTTAT 2600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 851 TyrAlaLysAlaMetValHisAspAsnPheAsnValLeuGluThrLeuMe 867

2601 GTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTA 2650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 867 tSerMetProArgAlaPheIleArgLysValProGlySerValValValT 884

2651 CCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTT 2700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 884 hrIleCysThrSerGlyAlaSerAspArgLeuGluLeuArgGlyAlaPhe 900

2701 GATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCG 2750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 901 AspIleSerLysGluThrPheGlyArgLysLeuLysAsnSerArgLeuAr 917

2751 CGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAA 2800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 917 gValPheSerArgAlaIleValGluAspSerIleLysValMetLysAlaM 934

2801 TGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATAT 2850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 934 etLysThrGluAspGlyLysProLeuProIleThrGluAspSerValTyr 950

2851 GCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCT 2900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 951 AlaPheIleMetGlyAsnValSerAsnValHisCysThrArgAlaGlyLe 967

2901 TCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGG 2950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 967 uLeuGlyGlySerLysAlaThrValValSerSerValSerLysGlyLeuV 984

2951 TAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTT 3000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 984 alAlaArgGlyAlaAlaThrLysAlaPheSerGlyIleThrSerPhePhe 1000

3001 TCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCT 3050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 SerThrGlySerLeuPheTyrAspArgGlyLeuThrGluAspGluArgLe 1017

3051 TGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATAC 3100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1017 uAspAlaLeuValArgThrGluAsnAlaIleAsnSerProValGlyIleL 1034

3101 TGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAA 3150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1034 euGluThrSerArgValAlaValSerLysValValAlaGlyThrLysGlu 1050

3151 TTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAA 3200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 PheTrpSerGluValSerLeuAsnAspPheThrThrPheValLeuArgAs 1067

3201 TAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTG 3250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1067 nLysValLeuIleGlyIlePheValAlaSerLeuGlyAlaAlaProIleA 1084

3251 CATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGC 3300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1084 laTrpLysTyrArgArgGlyIleAlaAlaAsnAlaArgArgTyrAlaGly 1100

3301 AGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTT 3350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 SerSerTyrGluThrLeuSerSerLeuSerSerGlnAlaAlaGlyGlyLe 1117

3351 ACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAA 3400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1117 uArgGlyLeuThrSerSerThrValSerGlyGlySerLeuValValArgA 1134

3401 GAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGT 3450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
1134 rgGlyPheSerSerAlaValThrValThrArgAlaThrValAlaLysArg 1150

3451 CAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGG 3500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 GlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAlaIleSerGl 1167

3501 CTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACTTAATGT 3550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1167 yCysSerMetLeuGlyIleTrpAlaHisAlaLeuProArgHisLeuMetP 1184

3551 TTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAATACT 3600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1184 hePhePheGlyLeuGlyThrLeuLeuGlyAlaArgAlaSerAlaAsnThr 1200

3601 TGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAGGTTGT 3650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 TrpLysPheGlyGlyPheSerAsnAsnTrpCysAlaValProGluValVa 1217

3651 TTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCTAGGGG 3700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1217 lTrpArgGlyLysSerValSerSerLeuLeuLeuProIleThrLeuGlyV 1234

3701 TATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCTCAACTTGCT 3750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1234 alSerLeuIleIleArgGlyLeuLeuAsnAspThrIleProGlnLeuAla 1250

3751 TACGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAGACGCTTAGGTA 3800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 TyrValProProValGluGlyArgAsnValTyrAspGluThrLeuArgTy 1267

3801 TTACCGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGC 3850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1267 rTyrArgAspPheAspTyrAspGluGlyAlaGlyProSerGlyThrGlnH 1284

3851 ATGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCTAGTGTCTCA 3900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1284 isGluAlaValProGlyAspAspAsnAspGlySerThrSerSerValSer 1300

3901 AGCTATGATGTTGTCACAAATGTGCGCGACGTGGGGATTAGCACCAACGG 3950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 SerTyrAspValValThrAsnValArgAspValGlyIleSerThrAsnGl 1317

3951 GGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACA 4000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1317 yGluValThrGlyGluGluGluThrHisSerProArgSerValGlnTyrT 1334

4001 CTTATGTCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAA 4050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1334 hrTyrValGluGluGluValAlaProSerAlaAlaValAlaGluArgGln 1350

4051 GGTGATCCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAG 4100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 GlyAspProSerGlySerGlyThrAlaAspAlaMetAlaPheValGluSe 1367

4101 TGTGAAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAA 4150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1367 rValLysLysGlyValAspAspValPheHisGlnGlnSerSerGlyGluT 1384

4151 CGGCTCGTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTC 4200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1384 hrAlaArgGluValGluValAspGlyLysGlyLeuLeuProGluSerVal 1400

4201 GTCGGTGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACAC 4250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 ValGlyGluAlaProThrGlnGluArgGlyArgAlaAlaAspGlyAsnTh 1417

4251 AGCACAAACCGCCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTC 4300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1417 rAlaGlnThrAlaValAsnGluGlyAspArgGluProValGlnSerSerL 1434

4301 TTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTA 4350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1434 euValSerSerProGlnAlaAspIleProLysValThrGlnSerGluVal 1450

4351 CATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGG 4400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 HisAlaGlnLysGluValLysGlnGluValProLeuAlaThrValSerGl 1467

4401 CGCCACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTG 4450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1467 yAlaThrProIleValAspGluLysProAlaProSerValThrThrArgG 1484

4451 GTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAA 4500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1484 lyValLysIleIleAspLysGlyLysAlaValAlaHisValAlaGluLys 1500

4501 AAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGA 4550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 LysGlnValGlnValGluGlnProLysGlnArgSerLeuThrIleAsnGl 1517

4551 AGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTG 4600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1517 uGlyLysAlaGlyLysGlnLeuCysMetPheArgThrCysSerCysGlyV 1534

4601 TGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAAC 4650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1534 alGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsn 1550

4651 GCATTTACCTTTGTCGATAACTTGAAAGGGAGCAGTGCGGTCTTTTTCTC 4700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 AlaPheThrPheValAspAsnLeuLysGlyArgSerAlaValPhePheSe 1567

4701 AAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAG 4750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1567 rLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerG 1584

4751 GGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGC 4800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1584 lyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSer 1600

4801 GTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACC 4850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 ValPheAspHisCysLeuValGlnLysTyrLysMetGlyGlyGlyValPr 1617

4851 ATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGA 4900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1617 oPheHisAlaAspAspGluGluCysTyrProSerAspAsnProIleLeuT 1634

4901 CGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGT 4950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1634 hrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysArgLysGly 1650

4951 GGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCC 5000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 GlyLysValMetValIleAsnValAlaSerGlyAspTyrPheLeuMetPr 1667

5001 TTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAG 5050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1667 oCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIleAspGluG 1684

5051 GGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGC 5100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1684 lyArgIleSerLeuThrPheArgAlaThrArgArgValPheGlyValGly 1700

5101 AGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGT 5150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 ArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSerProGlyVa 1717

5151 TCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAA 5200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1717 lProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIleThrProL 1734

5201 AATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGG 5250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1734 ysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluValLysGly 1750
```

```
5251 AGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 ArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArgLysCysGl 1767

5301 GTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCC 5350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1767 uTrpLeuArgAlaAspAsnProValMetAlaLeuGluProAspTyrThrP 1784

5351 CAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTG 5400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1784 roMetThrPheGluValValLysThrGlyThrSerGluAspAlaValVal 1800

5401 GAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTT 5450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 GluTyrLeuLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLe 1817

5451 GCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAG 5500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1817 uLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyValLeuLysV 1834

5501 TACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCG 5550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1834 alProAsnGlnValTyrGluSerLeuProGlyPheHisValTyrLysSer 1850

5551 GGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGA 5600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 GlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAs 1867

5601 CCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAG 5650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1867 pLeuProTyrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysA 1884

5651 ATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGAT 5700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1884 spValAspAlaValValAlaLeuGlyAspAsnLeuPheValCysAspAsp 1900

5701 ATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGC 5750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1901 IleLeuValPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAl 1917

5751 TCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCT 5800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1917 aArgCysGlyMetValGlyGluSerPheLysSerPheGluTyrLysCysT 1934

5801 ATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTC 5850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1934 yrAsnAlaProProGlyGlyGlyLysThrThrThrLeuValAspGluPhe 1950

5851 GTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTC 5900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1951 ValLysSerProAsnSerThrAlaThrIleThrAlaAsnValGlySerSe 1967

5901 TGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTC 5950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1967 rGluAspIleAsnMetAlaValLysLysArgAspProAsnLeuGluGlyL 1984

5951 TCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGG 6000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1984 euAsnSerAlaThrThrValAsnSerArgValValAsnPheIleValArg 2000

6001 GGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGG 6050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2001 GlyMetTyrLysArgValLeuValAspGluValHisMetMetHisGlnGl 2017

6051 CTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTT 6100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2017 yLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPheP 2034

6101 TTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGG 6150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2034 heGlyAspIleAsnGlnIleProPheIleAsnArgGluLysValPheArg 2050

6151 ATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTC 6200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2051 MetAspCysAlaValPheValProLysLysGluSerValValTyrThrSe 2067

6201 TAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGA 6250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2067 rLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSerSerMetT 2084

6251 CCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAG 6300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2084 hrValArgGlyThrGluLysCysTyrProGluLysValValSerGlyLys 2100

6301 GACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGA 6350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2101 AspLysProValValArgSerLeuSerLysArgProIleGlyThrThrAs 2117

6351 TGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGG 6400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2117 pAspValAlaGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuG 2134

6401 AGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCA 6450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2134 luLysSerAspMetLysArgSerLeuLysGlyLysGlyLysGluThrPro 2150

6451 GTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATT 6500
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2151 ValMetThrValHisGluAlaGlnGlyLysThrPheSerAspValValLe 2167

6501 GTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATA 6550
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2167 uPheArgThrLysLysAlaAspAspSerLeuPheThrLysGlnProHisI 2184

6551 TACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTG 6600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2184 leLeuValGlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeu 2200

6601 AGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACC 6650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2201 SerSerLysLeuAspAspLysValGlyThrTyrIleSerAspAlaSerPr 2217

6651 TCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGCTGGTTGCT 6700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
2217 oGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysP 2234

6701 TTCGAGGTATA 6711
     ||||||||||| 
2234 heArgGlyIle 2237
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:W21636

seq_documentation_block:

```
ID   W21636 standard; Protein; 1390 AA.
XX
AC   W21636;
XX
DT   19-SEP-1997  (first entry)
XX
DE   Grapevine leafroll virus helicase.
XX
KW   GLRaV; grapevine; Vitis; rootstock; leafroll; disease resistance;
KW   transgenic plant; tristeza virus; citrus; helicase.
XX
OS   Grapevine leafroll associated virus type 3 isolate NY1.
XX
PN   WO9722700-A2.
XX
PD   26-JUN-1997.
XX
PF   20-DEC-1996;   96WO-US20747.
XX
PR   21-DEC-1995;   95US-0009008.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
```

```
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 1997-341691/31.
DR   N-PSDB; T72214.
XX
PT   DNA encoding grape-vine leaf-roll virus proteins - useful to impart
PT   viral-resistance to Vitis scion or root-stock cultivar(s)
XX
PS   Claim 4; Page 69-75; 172pp; English.
XX
CC   A helicase (W21636) was identified as the expression product of
CC   open reading frame 1a (T72215) of a double-stranded cDNA molecule
CC   derived from grapevine leafroll associated virus type 3 (GLRaV-3)
CC   isolate NY-1 dsRNA.  The helicase, and other GLRaV-3 polypeptides
CC   (see also W21637-47), can be expressed in transgenic plants as a
CC   means of imparting GLRaV resistance to Vitis scion or rootstock
CC   cultivars or tristeza virus resistance to citrus scion or rootstock
CC   cultivars, or used to raise antibodies useful for detection of
CC   GLRaV.
XX
SQ   Sequence   1390 AA;


alignment_scores:
            Quality: 5247.00              Length:   1423
              Ratio:  4.214                Gaps:     17
 Percent Similarity:  87.491   Percent Identity: 74.139

alignment_block:
 US-09-301-906-3 x W21636   ..

 Align seg 1/1  to: W21636  from: 1  to: 1390

    2542 GTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGA 2591
         ||||||:::|||||||||:::::::::::|||||||||||||:::|||||
       1 ValSerThrTyrAlaLysSerValMetAsnAspAsnPheAsnIleLeuGl 17

    2592 GACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTG 2641
         |||||||:::::::::|||:::::|||||   ||||||||||||||||
      17 uThrLeuValThrLeuProLysSerPheIleValLysValProGlySerV 34

    2642 TTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGG 2691
         ||:::|||:::|||   |||||||||   ||||||:::|||||||||||
      34 alLeuValSerIleThrThrSerGlyIleSerAspLysLeuGluLeuArg 50

    2692 GGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAG 2741
         ||||||||||||:::||||||:::::::|||:::|||:::|||::::::||
      51 GlyAlaPheAspValSerLysLysAsnPheSerArgArgLeuArgSerSe 67

    2742 TCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCA 2791
         ||||||||||||||||||||||||||||||||||:::|||||||||||
      67 rArgLeuArgValPheSerArgAlaIleValGluAspThrIleLysValM 84

    2792 TGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGAT 2841
         |||||:::|||||||:::|||||||||||||||||||||||||:::||||||
      84 etLysGlyMetLysSerGluAspGlyLysProLeuProIleAlaGluAsp 100

    2842 TCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAG 2891
         ||||||||||||||:::   ||||||:::|||||||||||||||||||||
     101 SerValTyrAlaPheMetThrGlyAsnMetSerAsnValHisCysThrAr 117

    2892 GGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTA 2941
         |||||||||||||||||||||||||||   :::::::|||   :::   |
     117 gAlaGlyLeuLeuGlyGlySerLysAlaCysAlaAlaSerLeuAlaValL 134

    2942 AGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACG 2991
         |||||   ::::::|||:::::::::|||||   ||||||||||:::|||
     134 ysGlyAlaAlaSerArgAlaThrGlyThrLysLeuPheSerGlyLeuThr 150

    2992 TCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGA 3041
         ||||||:::|||:::|||:::|||||||||||:::|||||||||
     151 SerPheLeuSerAlaGlyGlyLeuPheTyrAspGluGlyLeuThrProGl 167

    3042 TGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGG 3091
         |||||||||||||||||||:::|||   |||:::|||:::||||||||||
     167 yGluArgLeuAspAlaLeuThrArgArgGluHisAlaValAsnSerProV 184

    3092 TGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGA 3141
         |||||:::|||||   :::      ::::::::::::||||||:::|||
     184 alGlyLeuLeuGluProGlyAlaSerValAlaLysArgValValSerGly 200

    3142 ACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGT 3191
         ||||||   |||   ||||||:::||||||:::|||||||||||||||||
     201 ThrLysAlaPheLeuSerGluLeuSerLeuGluAspPheThrThrPheVa 217

    3192 ATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGG 3241
         |:::::|||:::|||||||||||||:::|||:::   |||:::::   :
     217 lIleLysAsnArgValLeuIleGlyValPheThrLeuSerMetAlaLeuT 234

    3242 CCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGG 3291
         ::|||:::::|||||||||||||||:::||||||   ::::::
     234 hrProValValTrpLysTyrArgArgAsnIleAlaArgThrGlyValAsp 250

    3292 TACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGC 3341
                      :::   :::      ::::::::::::   |||:::
     251 ValPheHisArgAlaArgSerGlyThrAlaAlaIleGlyLeuGlnCysLe 267

    3342 CGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAG 3391
         :::|||   |||:::|||:::::::::::::::   |||:::      |
     267 uSerGlyGlyArgSerLeuAlaGlyAspAlaAlaArgGlyAlaLeuThrV 284

    3392 TCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTA 3441
         ||:::|||   |||:::||||||||||||:::|||||||||   ||||||
     284 alThrArgGlyGlyLeuSerSerAlaValAlaValThrArgAsnThrVal 300

    3442 GCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGC 3491
         |||:::|||||||||||||||||||||||||||||||||||||||||||
     301 AlaArgArgGlnValProLeuAlaLeuLeuSerPheSerThrSerTyrAl 317

    3492 CATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGC 3541
         |:::|||||||||:::::|||||||||||||||||||||||||||||||
     317 aValSerGlyCysThrLeuLeuGlyIleTrpAlaHisAlaLeuProArgH 334

    3542 ACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGC 3591
         |||||||||||||||||||||||||||||:::|||:::   ||||||
     334 isLeuMetPhePhePheGlyLeuGlyThrLeuPheGlyValSerAlaSer 350

    3592 GCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCC 3641
         :::|||:::|||::::::|||||||:::::|||:::      :::|||||
     351 ThrAsnSerTrpSerLeuGlyGlyTyrThrAsnSerLeuPheThrValPr 367

    3642 CGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTA 3691
         ||||:::::|||:::|||:::|||      |||||||||         :
     367 oGluLeuThrTrpGluGlyArgSerTyrArgSerLeuLeuProGlnAlaA 384

    3692 CGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCT 3741
         ::|||||:::|||||:::::|||||||||||||:::::|||:::|||
     384 laLeuGlyIleSerLeuValValArgGlyLeuLeuSerGluThrValPro 400

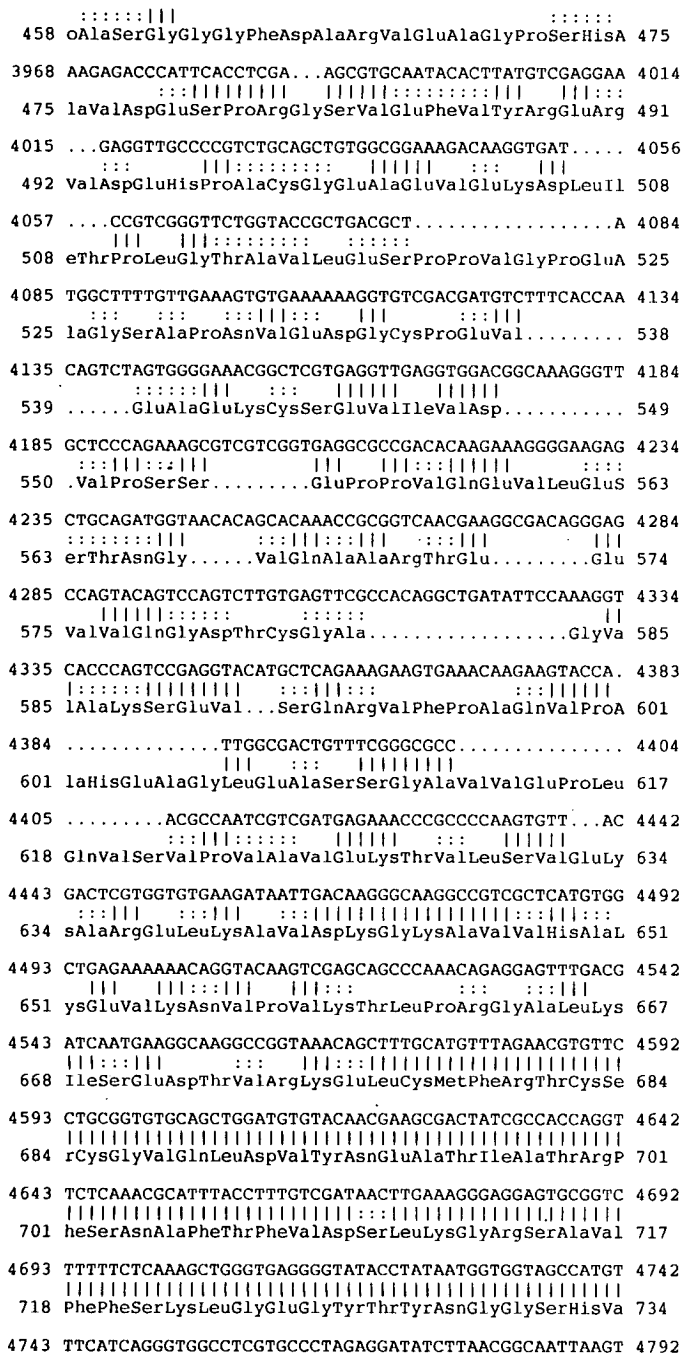
    3742 CAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAGAC 3791
         ||||||:::|||||||||||:::|||||||||||||||||||||:::::
     401 GlnLeuThrTyrValProProIleGluGlyArgAsnValTyrAspGlnAl 417

    3792 GCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTG 3841
         :|||:::::|||||||||||||||||||||:::|||||||||||!|||
     417 aLeuAsnPheTyrArgAspPheAspTyrAspAspGlyAlaGlyProSerG 434

    3842 GGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCT 3891
         |||||      ::::::   ||||||      |||   ::::::||||||
     434 lyThrAlaGlyGlnSerAspProGly...ThrAsnThrSerAspThrSer 449

    3892 AGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTGGGGATTAG 3941
         ||||||   |||                     |||   |||:::
     450 SerValPheSer.....................AspAspGlyLeuPr 458

    3942 CACCAACGGG.....................GAAGTTACTGGTGAAG 3967
```


458 oAlaSerGlyGlyGlyPheAspAlaArgValGluAlaGlyProSerHisA 475
3968 AAGAGACCCATTCACCTCGA...AGCGTGCAATACACTTATGTCGAGGAA 4014
475 laValAspGluSerProArgGlySerValGluPheValTyrArgGluArg 491
4015 ...GAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGAT..... 4056
492 ValAspGluHisProAlaCysGlyGluAlaGluValGluLysAspLeuIl 508
4057 CCGTCGGGTTCTGGTACCGCTGACGCT.................A 4084
508 eThrProLeuGlyThrAlaValLeuGluSerProProValGlyProGluA 525
4085 TGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCAA 4134
525 laGlySerAlaProAsnValGluAspGlyCysProGluVal..... 538
4135 CAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCAAAGGGTT 4184
539 GluAlaGluLysCysSerGluValIleValAsp..... 549
4185 GCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGGGGAAGAG 4234
550 .ValProSerSer.....GluProProValGlnGluValLeuGluS 563
4235 CTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGACAGGGAG 4284
563 erThrAsnGly......ValGlnAlaAlaArgThrGlu.....Glu 574
4285 CCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGGT 4334
575 ValValGlnGlyAspThrCysGlyAla.................GlyVa 585
4335 CACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACCA. 4383
585 lAlaLysSerGluVal...SerGlnArgValPheProAlaGlnValProA 601
4384 TTGGCGACTGTTTCGGGCGCC.............. 4404
601 laHisGluAlaGlyLeuGluAlaSerSerGlyAlaValValGluProLeu 617
4405 ACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTT...AC 4442
618 GlnValSerValProValAlaValGluLysThrValLeuSerValGluLy 634
4443 GACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGG 4492
634 sAlaArgGluLeuLysAlaValAspLysGlyLysAlaValValHisAlaL 651
4493 CTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACG 4542
651 ysGluValLysAsnValProValLysThrLeuProArgGlyAlaLeuLys 667
4543 ATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTC 4592
668 IleSerGluAspThrValArgLysGluLeuCysMetPheArgThrCysSe 684
4593 CTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGT 4642
684 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 701
4643 TCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTC 4692
701 heSerAsnAlaPheThrPheValAspSerLeuLysGlyArgSerAlaVal 717
4693 TTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGT 4742
718 PhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisVa 734
4743 TTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGT 4792
734 lSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysT 751
4793 ACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGA 4842
751 yrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 767
4843 GGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCC 4892
768 GlyValProPheHisAlaAspAspGluGluCysTyrProSerAspAsnPr 784
4893 TATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCA 4942
784 oIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysA 801
4943 GGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTT 4992
801 rgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 817
4993 CTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCAT 5042
818 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIl 834
5043 CGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTG 5092
834 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArgValPheG 851
5093 GTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142
851 lyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSer 867
5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
868 ProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIl 884
5193 CACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAG 5242
884 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 901
5243 TCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGG 5292
901 alLysGlyArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArg 917
5293 AAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGA 5342
918 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuLysProGl 934
5343 CTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATG 5392
934 yTyrThrProMetThrPheGluValValLysAlaGlyThrSerGluAspA 951
5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC 5442
951 laValValGluTyrLeuLysTyrLeuAlaIleGlyIleGlyArgThrTyr 967
5443 AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGT 5492
968 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyVa 984
5493 TCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTT 5542
984 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1001
5543 ACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGT 5592
1001 yrLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArg 1017
5593 GTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAA 5642
1018 ValArgAspLeuProTyrValPheIleAlaGluLysGlyIlePheIleLy 1034
5643 GGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTAT 5692
1034 sGlyLysAspValAspAlaValValAlaLeuGlyAspAsnLeuSerValC 1051

```
5693 GCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTG 5742
     ||||||||||||||||||||||||||||||||||||:::|||||||||
1051 ysAspAspIleLeuValPheHisAspAlaIleAsnLeuMetGlyAlaLeu 1067

5743 AAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATA 5792
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1068 LysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTy 1084

5793 TAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGG 5842
     |||||||||||||||||||||||||||||||||||||    |||||||
1084 rLysCysTyrAsnAlaProProGlyGlyGlyLysThrThrMetLeuValA 1101

5843 ACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTG 5892
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 spGluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1117

5893 GGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTT 5942
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1118 GlySerSerGluAspIleAsnMetAlaValLysLysArgAspProAsnLe 1134

5943 GGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTA 5992
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1134 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPheI 1151

5993 TCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATG 6042
     ||||||||||||||||||||||||||||||||||||:::|||||||||
1151 leValArgGlyMetTyrLysArgValLeuValAspGluValTyrMetMet 1167

6043 CATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGG 6092
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1168 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGl 1184

6093 CCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGG 6142
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1184 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 1201

6143 TGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTA 6192
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 alPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 1217

6193 TACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTC 6242
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1218 TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 1234

6243 CTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTA 6292
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1234 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 1251

6293 GCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGA 6342
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 erGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 1267

6343 ACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGAC 6392
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1268 ThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 1284

6393 CCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAG 6442
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1284 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLysG 1301

6443 AAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGAT 6492
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 1317

6493 GTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACA 6542
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1318 ValValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGl 1334

6543 ACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATG 6592
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1334 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 1351

6593 CCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGAC 6642
     ||||||||||||||:::||||||||||||||||||||||||||||||||
1351 laAlaLeuSerSerGluLeuAspAspLysValGlyThrTyrIleSerAsp 1367

6643 GCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGC 6692
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1368 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 1384

6693 TGGTTGCTTTCGAGGTATA 6711
     |||||||||||||||||||
1384 aGlyCysPheArgGlyIle 1390
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58145

seq_documentation_block:

```
ID   Y58145 standard; Protein; 291 AA.
XX
AC   Y58145;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein helicase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49205.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 15; Fig 11; 84pp; English.
XX
CC   This sequence represents the helicase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
```

```
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   291 AA;


alignment_scores:
            Quality: 1495.00          Length:    291
              Ratio:  5.137             Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-301-906-3 x Y58145   ..

 Align seg 1/1  to: Y58145  from: 1  to: 291

    5764 GTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCC 5813
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ValGlyGluSerPheLysSerPheGluTyrLysCysTyrAsnAlaProPr 17

    5814 AGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCA 5863
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 oGlyGlyGlyLysThrThrThrLeuValAspGluPheValLysSerProA 34

    5864 ATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACATAAAT 5913
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 snSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspIleAsn 50

    5914 ATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTAC 5963
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 MetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSerAlaTh 67

    5964 CACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAA 6013
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rThrValAsnSerArgValValAsnPheIleValArgGlyMetTyrLysA 84

    6014 GGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTA 6063
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 rgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuGlnLeu 100

    6064 GGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAA 6113
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAspIleAs 117

    6114 TCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTG 6163
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 nGlnIleProPheIleAsnArgGluLysValPheArgMetAspCysAlaV 134

    6164 TTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAGG 6213
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 alPheValProLysLysGluSerValValTyrThrSerLysSerTyrArg 150

    6214 TGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAAC 6263
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 CysProLeuAspValCysTyrLeuLeuSerSerMetThrValArgGlyTh 167

    6264 GGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAG 6313
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 rGluLysCysTyrProGluLysValValSerGlyLysAspLysProValV 184

    6314 TAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAA 6363
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 alArgSerLeuSerLysArgProIleGlyThrThrAspAspValAlaGlu 200

    6364 ATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATAT 6413
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 IleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSerAspMe 217

    6414 GAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGC 6463
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 tLysArgSerLeuLysGlyLysGlyLysGluThrProValMetThrValH 234

    6464 ATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAG 6513
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 isGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgThrLys 250

    6514 AAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTT 6563
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuValGlyLe 267

    6564 GTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGG 6613
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 uSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLysLeuA 284

    6614 ACGATAAGGTCGGCACATATATT 6636
         |||||||||||||||||||||||
     284 spAspLysValGlyThrTyrIle 291

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58144

seq_documentation_block:
ID   Y58144 standard; Protein; 272 AA.
XX
AC   Y58144;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein methyltransferase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49204.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 5; Fig 8; 84pp; English.
XX
CC   This sequence represents the methyltransferase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
```

```
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   272 AA;


alignment_scores:
            Quality: 1432.00              Length:    272
              Ratio:  5.265                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-301-906-3 x Y58144   ..

 Align seq 1/1  to: Y58144  from: 1  to: 272

    1378 CTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTC 1427
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 LeuLysProArgGluArgGluLysLeuArgGluLeuProGluLeuSe 17

    1428 GATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCA 1477
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rIleGlnPheSerAspSerValArgSerSerHisProPheAlaAsnAlaM 34

    1478 TGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGC 1527
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 etArgSerCysPheAsnGlyIlePheSerArgArgCysGlyAsnValCys 50

    1528 TTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGT 1577
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 PhePheAspIleGlyGlySerPheThrTyrHisValLysAlaGlyHisVa 67

    1578 GAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGA 1627
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lAsnCysHisValCysAsnProValLeuAspValLysAspValLysArgA 84

    1628 GAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGTG 1677
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 rgIleAsnGluIleLeuPheLeuSerThrAlaGlyGlyAspSerTyrVal 100

    1678 TCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTGTAG 1727
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 SerSerAspLeuLeuThrGluAlaAlaSerLysSerValSerTyrCysSe 117

    1728 TCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTTATGGTGG 1777
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 rArgGluSerGlnAsnCysAspSerArgAlaAspAlaGlyPheMetValA 134

    1778 ATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAG 1827
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 spValTyrAspIleSerProGlnGlnValAlaGluAlaMetAspLysLys 150

    1828 GGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTA 1877
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GlyAlaLeuValPheAspIleAlaLeuMetPheProValGluLeuLeuTy 167

    1878 CGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGG 1927
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 rGlyAsnGlyGluValTyrLeuGluGluLeuAspThrLeuValLysArgG 184

    1928 AAGGTGATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAA 1977
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 luGlyAspTyrLeuAlaTyrAsnValGlyGlnCysGlyGluMetTyrGlu 200

    1978 CATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGCAC 2027
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 HisSerPheSerAsnValSerGlyPhePheThrPheSerTyrValArgTh 217

    2028 TTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTT 2077
         ||||||||||||||||||||||||||||||||||||||||||||||||||

     217 rSerSerGlyAsnValPheLysLeuGluTyrGluGlyTyrArgCysGlyT 234

    2078 ACCATCATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACTGAGGTT 2127
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 yrHisHisLeuThrMetCysArgAlaGlnLysSerProGlyThrGluVal 250

    2128 ACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCAT 2177
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ThrTyrArgSerLeuValProSerPheValGlyLysSerLeuValPheIl 267

    2178 ACCTGTTGTAGCTGGT 2193
         ||||||||||||||||
     267 eProValValAlaGly 272

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W73476

seq_documentation_block:
ID   W73476 standard; Protein; 2639 AA.
XX
AC   W73476;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Grapevine leafroll virus type 2 (GRLaV-2) polyprotein.
XX
KW   GRLaV-2; closterovirus; grape; tobacco; transgenic plant;
KW   disease resistance; virus resistance; beet yellows virus;
KW   tristeza virus; protease; methyltransferase; helicase.
XX
OS   Grapevine leafroll virus type 2.
XX
PN   WO9853055-A1.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998;   98WO-US10313.
XX
PR   20-MAY-1997;   97US-0047194.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K,  Zhu H;
XX
DR   WPI; 1999-045307/04.
DR   N-PSDB; V08864 and V08874.
XX
PT   Grapevine leafroll virus (type 2) proteins and polypeptides - and
PT   encoding DNA, useful e.g. to impart grapevine leafroll resistance to
PT   grape and tobacco plants and detect grapevine leafroll virus
XX
PS   Claim 4; Page 25-33; 151pp; English.
XX
CC   This is the amino acid sequence of a 294 kDa polyprotein encoded by
CC   open reading frame ORF1a (see V08864) of grapevine leafroll virus
CC   type 2 (GLRaV-2) RNA (see V08874).  It contains conserved domains
CC   characteristic of 2 papain-like proteases, a methyltransferase and
CC   a helicase.  The GLTaV-2 genome includes 9 open reading frames
CC   (see V08864-72) encoding the polyprotein, an RNA-dependent RNA
CC   polymerase, heat shock proteins, coat proteins and other proteins
CC   of unknown function (see W73476-84).  These can be used to produce
CC   antibodies useful to detect GLRaV in samples e.g. by ELISA
CC   (claimed).  The nucleic acid molecules can be used to produce
CC   probes and primers for such detection, and to transform host cells
CC   (especially Agrobacterium vitis, Agrobacterium tumefaciens, grape,
CC   citrus, beet or tobacco cells) and produce transgenic plants
CC   (claimed).  They can be used to impart GLRaV-2 resistance to Vitis
CC   scion or rootstock cultivars or Nicotiana cultivars (claimed).
CC   Because extensive similarity exists between hsp70-related sequence
CC   regions of GLRaV-2 and other closteroviruses, the DNA may also be
CC   used to impart beet yellows virus resistance to beet cultivars or
CC   tristeza virus resistance to citrus scion cultivar/rootstock
CC   cultivars (claimed).
XX
```

SQ Sequence 2639 AA;

alignment_scores:
Quality: 802.50 Length: 2475
Ratio: 0.681 Gaps: 115
Percent Similarity: 47.596 Percent Identity: 22.424

alignment_block:
US-09-301-906-3 x W73476 ..

Align seg 1/1 to: W73476 from: 1 to: 2639

```
 454 CAAGATGAGCAGAAGCCCAAACGGCAAGCGTCACATTGGGCTGTTAAGCC 503
     |||::::::   |||   ||||||:::   :::|||   :::|||:::
 324 GlnSerServValLysThrLysArgGluAspAlaHisArgThrValGluGl 340

 504 AACAGCTGTTGGCGTCCACGTACCACTTCCTAAAAAACAGGAAGCACTGG 553
         |||:::|||      |||      |||:::::::::::::      |
 340 uArgAlaAlaGlyGlySerValGluGlnProArgGlnLysArgIleAspG 357

 554 AGCCA...GCGCAATCAGTCCCACAACAGTCGTTG..............  585
     ||      :::      ||||||:::   ::::::
 357 luLysGlyCysGlyArgValProSerGlyGlyPheSerHisLeuLeuVal 373

 586 ............GAGGAGAAGGCCGCCTTGACGTTTGGCCTTTTCTTCAG 623
                 |||   :::     ::::::   ||||||:::
 374 GlyAsnLeuAsnGluValArgArgLysValAlaAlaGlyLeuLeuArgPh 390

 624 TAAAGGTGGGGGTGATGAGAGCGACGCTGTCATCTTGCGGAAAGGGAAAT 673
      :::   |||||||||   :::
 390 eArgValGlyGlyAspMetAsp............................ 397

 674 TGTTTAACAGGGCCCTTAATGTTCCTATTGATGTAAAGAACACGTTCGTT 723
       |||:::|||:::::::::::::                 :::   :::|||
 398 ..PheHisArgSerPheSerThrGlnAlaGly...HisArgLeuLeuVal 412

 724 TGGGCTAAA.......................ATCTGGGATGAAGCCTC 749
     |||   :::                       :::::::::   :::::
 413 TrpArgArgSerSerArgSerValCysLeuGluLeuTyrSerProSerLy 429

 750 TCGTAGGAGAGGGTATTTTTACGTCAAA.........GATAGAGCTGTTA 790
     ::::         |||       :::              |||   |||:::
 429 sAsnPheLeuArgTyrAspValLeuProCysSerGlyAspTyrAlaAlaM 446

 791 AATTCTTCCCTATTGTGCGGGGTAGGGCTACGATCGAGGACTTCATCGTG 840
       |||         :::   ||||||      :::
 446 etPheSerPheAlaAlaGlyGlyArgPheProLeu............... 457

 841 AATACAGCCCCAGGGTGTGATGTTGCCTTGCCGCGCATTGAGTTGTGGAG 890
                       |||   :::   ||||||:::
 458 .....................ValLeuMetThrArgIleArgTyr..... 465

 891 TATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTA 940
                                   :::|||:::|||::::::
 466 ............................ProAsnGlyPheCysTyrLeuA 473

 941 ACAATGAGAGGCTGAGGGGAGAAATTTACAGACGTCGTTGCTTCTCATCT 990
       :::   |||           :::      |||           :::
 473 laHisCysArgTyrAlaCysAlaPheLeuLeuArgGlyPheAspProLys 489

 991 TCCTTTTCGATAGGT.................................TT 1007
       |||:::||||||                                  ::
 490 ArgPheAspIleGlyAlaPheProThrAlaAlaLysLeuArgAsnArgMe 506

1008 CTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTTGCGGGCA 1057
     ::::   :::||||||   |||||||||
 506 tValSerGluLeuGlyGluArgSerLeu..................... 515

1058 CGAACATACTACACATGCCATCACTCAATGAAGAGCGTACCTTTGGGTGG 1107
                      :::||||||         ::::::
 516 ..................GlyLeuAsnLeuTyrGlyAlaTyrThrSer 525

1108 AAGGGC..............................GGAGACGTCTA 1124
     :::|||                                 |||:::
 526 ArgGlyValPheHisCysAspTyrAspAlaLysPheIleLysAspLeuAr 542

1125 TCTCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGT 1174
      |||              :::|||::::::||||||  .:::      :
 542 gLeu............MetSerAlaValIleAlaGly...LysAspGlyV 554

1175 TGGGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAG 1224
     ::         :::   ::::::::::::   |||::::::|||:::
 554 alGluGluValValProSerAspIleThrProAlaMetLysGlnLys... 569

1225 GTCTATTCATCGGTCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGA 1274
                ::::::::::::::::   :::||||||
 570 ............ThrIleGluAlaValTyrAspArgLeuTyrGlyGlyTh 582

1275 CCAATCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTT 1324
      :::|||   |||         ::::::      |||   |||:::   :
 582 rAspSerLeuLeuLysLeuSerIleGluLysAspLeuIleAspPheLysA 599

1325 CTCAAAGGGACGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTG 1374
     :::::   :::::::::   ::::::::::|||      :::
 599 snAspValGlnSerLeuLysLysAspArgProIleValLysValProPhe 615

1375 TTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCT 1424
     :::::::::          :::::::::|||   ::::::::::|||:::::
 616 TyrMetSerGluAlaThrGlnAsnSerLeuThrArgPheTyrProGlnPh 632

1425 TTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATG 1474
     ::::::::::||||||   |||   :::|||:::||||||   |||   |
 632 eGluLeuLysPheSerHisSerSerHisSerAspHisProAlaAlaAlaA 649

1475 CCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTG 1524
     ||   |||         |||      :::   |||   |||||||||
 649 laSerArgLeuLeuGluAsnGluThrLeuValArgLeuCysGlyAsn... 664

1525 TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCT...GG 1571
            ||||||||||||      :::|||:::   :::
 665 SerValSerAspIleGlyGlyCysProLeuPheHisLeuHisSerLysTh 681

1572 CCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGA 1621
      :::   :::   ||||||||||:::||||||||||||   ||||||::::
 681 rGlnArgArgValHisValCysArgProValLeuAspGlyLysAspAlaG 698

1622 AGCGGAGA...ATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGAT 1668
     ::||||||   ::::::::::::   :::   ::::::      ||||||
 698 lnArgArgValValArgAspLeuGlnTyrSerAsnValArgLeuGlyAsp 714

1669 TCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTC 1718
               :::|||   :::   |||   :::   ::::::::::
 715 ............AspAspLysIleLeuGlu...GlyProArgAsnIleAs 726

1719 TTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTT 1768
     :   |||                    ||||||   ::::::::::|||   :
 726 pIleCysHisTyrProLeuGlyAlaCysAspHisGluSerSerAlaMetM 743

1769 TTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATG 1818
     ::||||||:::|||||||||   |||      :::::::::   ||||||
 743 etMetValGlnValTyrAspAlaSerLeuTyrGluIleCysGlyAlaMet 759

1819 GATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGA 1868
       ||||||   :::   ::::      ::::::::::::   |||   ||
 760 IleLysLysLysSerArgIleThrTyrLeuThrMetValThrProGlyGl 776

1869 GTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGG 1918
     |:::|||   |||:::      ||||||:::|||:::||||||      :
 776 uPheLeuAspGlyArgGluCysValTyrMetGluSerLeuAspCysGluI 793

1919 TGAAGAGGGAA......GGTGATTACCTGGCCTACAATGTTGGTCAGTGT 1962
     :::::   :::      :::|||   :::   |||:::   |||
 793 leGluValAspValHisAlaAspValValMetTyrLysPheGly...... 807

1963 GGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTT 2012
```

```
      ::::::   |||:::|||::::::|||   ::::::   ::::::|||
  808 SerSerCysTyrSerHisLysLeuSerIleIleLysAspIleMetThrTh 824

 2013 TTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGG 2062
      |||:::   |||   :::|||   :::|||::::::|||
  824 rProTyrLeu...ThrLeuGlyGlyPheLeuPheSerValGluMetTyrG 840

 2063 GATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCA 2112
       |||   |||   :::::::::   :::   :::::::::   |||
  840 luValArgMetGlyValAsnTyrPheLysIleThrLysSerGluValSer 856

 2113 CCTGGAACTGAGGTTACG...........TATAGGTCGTTG........ 2142
      |||:::   :::   |||           ||||||
  857 ProSerIleSerCysThrLysLeuLeuArgTyrArgArgAlaAsnSerAs 873

 2143 ..............GTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCA 2176
                     :::|||   |||      |||      :::   :
  873 pValValLysValLysLeuProArgPheAspLysLysArgArgMetCysL 890

 2177 TACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCGAT 2226
      ::|||                    ::::::   ||||||   ||||||
  890 euPro.....................GlyTyrAspThrIleTyrLeuAsp 899

 2227 TCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGAC 2276
      |||   ||||||:::|||:::::::::|||::::::   :::   :::::
  900 SerLysPheValSerArgValPheAspTyrValValCysAsnCysSerAl 916

 2277 ATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAA 2326
      :   :::::::::||||||||::::::   :::   ::::::|||:::|
  916 aValAsnSerLysThrPheGluTrpValTrpSerPheIleLysSerSerL 933

 2327 AGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATT 2376
      ||::::::|||||   ::::::::::::::|||:::   |||::::::
  933 ysSerArgValIleIleSerGlyLysIleIleHisLysAspValAsnLeu 949

 2377 TCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGAT 2426
      :::        :::   ::::::::::::::::::::|||::::::::
  950 AspLeuLysTyrValGluSerPheAlaAlaValMetLeuAlaSerGlyVa 966

 2427 TAAGGATAGG.................GCGAAGAGTATTCGCTCCTATA 2458
      ::::::::|||                |||||::::::      ::::
  966 lArgSerArgLeuAlaSerGluTyrLeuAlaLysAsnLeuSerHisPheS 983

 2459 ACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTC 2508
      ::
  983 er.............................................. 983

 2509 TTTCAGACCGTAGGCGATTGT...TTTTCGAACGCAGTCTCCGTCTATGC 2555
                  |||||||||   |||   :::|||:::|||
  984 ............GlyAspCysSerPheIleGluAlaThrSerPheValLe 996

 2556 TAAGGCAATGGTGCACGAT.........AACTTCAACGTTTTGGAGACGC 2596
       :::      :::::::::         |||||||||      |||   |
  996 uArgGluLysIleArgAsnMetThrLeuAsnPheAsn......GluArgL 1011

 2597 TTATGTCTATGCCCAGA......GCGTTCATCCGTAAAGTACCTGGGTCT 2640
      ||:::::::::   :::      ||||||               :::
 1011 euLeuGlnLeuValLysArgValAlaPheAlaThrLeuAspValSerPhe 1027

 2641 GTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAG 2690
      :::   :::            ::::::   :::|||      |||   ::
 1028 LeuAspLeuAspSerThrLeuGluSerIleThrAspPheAlaGluCysLy 1044

 2691 GGGTGCCTTTGATATTTCGAAG.................GAGACCTTCG 2722
      :   |||:::::::::::::::                 |||:::   :
 1044 sValAlaIleGluLeuAspGluLeuGlyCysLeuArgAlaGluAlaGluA 1061

 2723 GTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTG 2772
      :::::|||::::::|||                              ::::::
 1061 snGluLysIleArgAsn.........................LeuAla 1068

 2773 GAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAA.........GA 2813
         |||||||||   :::      |||   :::               ||
```

```
 1069 GlyAspSerIleAlaAlaLysLeuAlaSerGluIleValValAspIleAs 1085

 2814 TGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGG 2863
      |:::||||||   |||                              :::|
 1085 pSerLysProSerProLysGln.......................ValG 1094

 2864 GGAACGTTTCTAACGTCCACTGT..............ACGAGGGCAGGT 2898
      |||||   |||:::   ::::::                  |||   |||
 1094 lyAsnSerSerSerGluAsnAlaAspLysArgGluValGlnArgProGly 1110

 2899 CTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTT 2948
      |||   |||||||||:::      ||||||::::::
 1111 LeuArgGlyGlySerArgAsnGlyValValGlyGlu.............. 1122

 2949 GGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCT 2998
                                          :::      |||
 1123 ..................................PheLeuHisPheV 1127

 2999 TTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGG 3048
               ::::::|||         :::      |||::::::::|||
 1127 alValAspSerAlaLeuArgLeuPheLysTyrAlaThrAspGlnGlnArg 1143

 3049 CTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCAT 3098
      :::   :::   ||||||            :::::::|||   |||::::::
 1144 IleLysSerTyrValArg.........PheLeuAspSerAlaValSerPh 1157

 3099 ACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAG 3148
      :|||:::   :::         :::|||   ::::::         :::|
 1157 eLeuAspTyrAsnTyrAspAsnLeuSerPheIleLeuArgValLeuSerG 1174

 3149 AATTTTGGAGT..................................... 3159
      ||   :::|||
 1174 luGlyTyrSerCysMetPheAlaPheLeuAlaAsnArgGlyAspLeuSer 1190

 3160 ..............................GAAGTTTCCTTAAATGA 3176
                                     ||||||:::   :::
 1191 SerArgValArgSerAlaValCysAlaValLysGluValAlaThrSerCy 1207

 3177 CTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGG 3226
         ::::::   |||   :::   ||||||:::|||
 1207 sAlaAsnAlaSerValSerLysAlaLysValMetIle............. 1219

 3227 CGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGC...GGAATT 3273
        :::::::::||||||      |||      ::::::      |||:::
 1220 ..ThrPheAlaAlaAlaValCysAlaMetMetPheAsnSerCysGlyPhe 1235

 3274 GCGGCTAACGCTAGAAGGTACGCGGGC..................... 3300
      ::::::::::::|||:::|||   :::
 1236 SerGlyAspGlyArgGluTyrLysSerTyrIleHisArgTyrThrGlnVa 1252

 3301 ............................AGTAGTTACGAAACTCTAAGCT 3322
                                   |||||||||      ::::::
 1252 lLeuPheAspThrIlePhePheGluAspSerSerTyrLeuProIleGluV 1269

 3323 CGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGC... 3369
        |||||||||      :::|||::::::      |||   ||||||
 1269 alLeuSerSerAlaIleCysGlyAlaIleValThrLeuPheSerSerGly 1285

 3370 .........ACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTC 3410
                  ::::::::::::      |||   :::   :::|||||||||
 1286 SerSerIleSerLeuAsnAlaPheLeuLeuGlnIleThrLysGlyPheSe 1302

 3411 GTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCT 3460
      |      |||:::|||   :::::::::            :::|||
 1302 rLeuGluValValValArgAsnValVal............ArgValThrH 1315

 3461 TAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTG....... 3503
        :::|||   :::   :::|||:::   ::::|||   |||
 1315 isGlyLeuSerThrThrAlaThrAspGlyValIleArgGlyValPheSer 1331

 3504 CAGTATGTTAGGCATTTG......GGCACATGCTCTTCCACGGCA..... 3542
      |||   |||   ||||||      |||:::   ::::::::::|||
 1332 GlnIleValSerHisLeuLeuValGlyAsnThrGlyAsnValAlaTyrGl 1348
```

```
3543 ..........................................CTTA 3546
                                                ::::
1348 nSerAlaPheIleAlaGlyValValProLeuLeuValLysLysCysValS 1365

3547 ATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAA 3596
     ::::::::::::   :::|||:::   :::::::::          ::::::
1365 erLeuIlePheIleLeuArgGluAspThrTyrSerGlyPheIleLysHis 1381

3597 TACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAGG 3646
        ::::::   :::   :::|||:::   :::::::::      :::||
1382 GlyIleSerGluPheSerPheLeuSerSerIleLeuLysPheLeuLysGl 1398

3647 TTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCTA 3696
     |   |||:::   |||       :::::::|||                :
1398 yLysLeuValAspGluLeuLysSerIleIle.................G 1409

3697 GGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCTCAACT 3746
     ::|||:::|||||||   :::       :::   ::::::   |||:::
1409 lnGlyValPheAspSerAsnLysHisValPheLysGluAlaThrGlnGlu 1425

3747 TGCTTACGTCCCACCGGTA.............................. 3765
     ::::::|||   ||| |||
1426 AlaIleArgThrThr.ValMetGlnValProValAlaValValAspAlaL 1442

3766 ...........GAGGGGAGGAATGTGTACGATGAGACGCTTAGGTATTAC 3804
                   |||:::          :::   |||   |||
1442 euLysSerAlaAlaGlyLysIleTyrAsnAsnPheThrSerArg...... 1456

3805 CGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGA 3854
     |||   |||      ||||||||||::::::   :::|||:::   :::||
1457 ArgThrPheGlyLysAspGluGlySerSerSerAspGlyAlaCysGluGl 1473

3855 AGCGGTTCCCGGTGACGATAACGATGGA...TCCACTTCTAGTGTCTCAA 3901
     |            |||::::::::::|||   :::   ::::::   ||||
1473 uTyrPheSerCysAspGluGlyGluGlyProGlyLeuLysGlyGlySerS 1490

3902 GCTATGATGTTGTCACAAATGTGC............GCGACGTGGGGATT 3939
     |||||    :::|||   ::::::                  ||||||
1490 erTyrGly.PheSerIleLeuAlaPhePheSerArgIleMetTrpGly.. 1505

3940 AGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAG 3989
        :::   :::|||:::||||||         :::      |||:
1506 .......AlaArgArgLeuIleValLys.........ValLysHisGluC 1517

3990 CGTGCAATACACTTATGTCGAGGAAGAGGTTGCCC............CGT 4027
     ::   :::   |||::::::          |||              |||
1517 ysPheGlyLysLeuPheGluPheLeuSerLeuLysLeuHisGluPheArg 1533

4028 CTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCT 4077
        ::::::::: ||||||:::|||   :::||||||   :::
1534 ThrArgValPhe.GlyLysAsnArgThrAspValGlyValTyrAspPheL 1550

4078 GACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTT 4127
          ||| :::||||||::::::
1550 euProThrGly.IleValGluThrLeu..................... 1558

4128 TCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCA 4177
               ||||||   |||  :::   ::::::|||
1559 .........SerSerIleGluGluCysAspGlnIleGlu.......... 1568

4178 AAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGG 4227
          ||||||   :::::::::   |||
1569 ..GluLeuLeuGlyAspAspLeuLysGly..................... 1577

4228 GGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGA 4277
                  |||   :::::::::   |||   :::|||      ::
1578 ............AspLysAspAlaSerLeuThrAspMetAsnTyrPheGl 1590

4278 C...AGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATA 4324
     :      |||        :::|||:::   :::   |||   |||   :
1590 uPheSerGluAspPheLeuAlaSerIleGluGluProProPheAlaGlyL 1607
```

```
4325 TTCCAAAGGTCACCCAGTCC............................. 4344
     ::      :::::::::
1607 euArgGlyGlySerLysAsnIleAlaIleLeuAlaIleLeuGluTyrAla 1623

4345 ............GAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACC 4382
                   ::::::   |||:::|||      |||:::   :::
1624 HisAsnLeuPheArgIleValAlaSerLysCysSerLysArgProLeuPh 1640

4383 ATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCCC 4432
      ||||||   ::::::   ::::::   ::::::::::::
1640 eLeuAlaPheAlaGluLeuSerSerAlaLeuIleGluLys......PheL 1655

4433 CAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTC 4482
       :::|||      |||                  |||::::::   |||
1655 ysGluValPheProArg..................LysSerGlnLeuVal 1665

4483 GCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAG 4532
     |||   |||   |||      |||   :::      :::   :::::::||
1666 AlaIleValArgGluTyrThrGln...ArgPheLeuArgSerArgMetAr 1681

4533 GAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTA 4582
     |:::|||   :::|||:::      ::: |||:::|||||||||   |||
1681 gAlaLeuGlyLeuAsnAsnGluPheVal.ValLysSerPheAlaAspLeu 1697

4583 GAACGTGTTCCTGCGGTG............TGCAGCTGGATGTGTACAAC 4620
           :::|||||||:::                |||  |||
1698 ......LeuProAlaLeuMetLysArgLysValSer.GlySerPheLeuA 1712

4621 GAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAA 4670
       |||   ||||||       :::              |||:::|||
1712 laSerValTyrArg...ProLeuArgGlyPheSerTyrMetCysValSer 1727

4671 CTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATA 4720
        ||||||:::|||      :::   |||   :::      |||:::
1728 AlaGluArgArgGluLysPhePheAlaLeuValCysLeuIleGlyLeuSe 1744

4721 CCTATAATGGTGGTAGCCATGTTTCATCAGG................... 4751
      |||            |||   |||:::|||
1744 rLeu............ProPhePheValArgIleValGlyAlaLysAlaC 1757

4752 ...........GTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAA 4790
                ||||:::|||:::   |||   |||
1757 ysGluGluLeuValSerSerAlaArgArgPheTyr.............. 1768

4791 GTACCCAAGCGTCTTCGA............................... 4808
        :::|||::::::
1769 ......GluArgIleLysIlePheLeuArgGlnLysTyrValSerLeuSe 1783

4809 .............CCACTGTTTAGTGCAGAAGTACAAGAT.......... 4835
                  |||||||||:::::::|||:::|||
1783 rAsnPhePheCysHisLeuPheSerSerAspValAspAspSerSerAlaS 1800

4835 ..................................................... 4835

1800 erAlaGlyLeuLysGlyGlyAlaSerArgMetThrLeuPheHisLeuLeu 1816

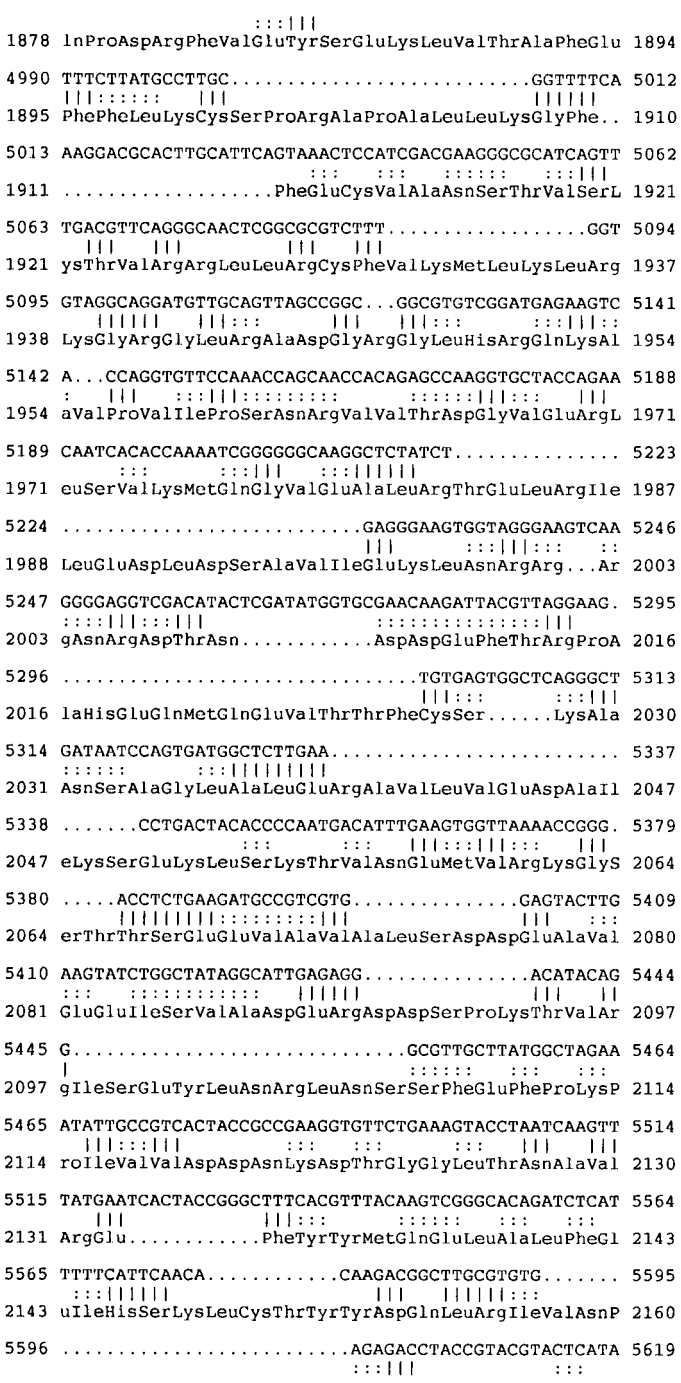
4836 ............................GGGTGGAGGCGTACCATTCC 4855
                                |||||||:::
1817 ValArgLeuAlaSerAlaLeuLeuSerLeuGlyTrpGluGlyLeu..... 1831

4856 ACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATC...TTGACG 4902
               :::   :::||||||       |||   :::   |||
1832 ..........LysLeuLeuLeuSer.HisHisAsnLeuLeuPheLeuCys 1844

4903 GTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGG 4952
           ||||||      :::|||         |||         |||
1845 PheAlaLeuValAspAspValAsnValLeuIleLysValLeuGlyGlyLe 1861

4953 TAAGGTCATGGTCATAAACGTAGCTTCG...................... 4980
      :::   :::|||      :::   |||
1861 uSerPhePheValGlnProIlePheSerLeuPheAlaAlaMetLeuLeuG 1878

4981 .............GGTGACTAT.......................... 4989
```


1878 lnProAspArgPheValGluTyrSerGluLysLeuValThrAlaPheGlu 1894
4990 TTTCTTATGCCTTGC.....GGTTTTCA 5012
1895 PhePheLeuLysCysSerProArgAlaProAlaLeuLeuLysGlyPhe.. 1910
5013 AAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTT 5062
1911 PheGluCysValAlaAsnSerThrValSerL 1921
5063 TGACGTTCAGGGCAACTCGGCGCGTCTTT.....GGT 5094
1921 ysThrValArgArgLeuLeuArgCysPheValLysMetLeuLysLeuArg 1937
5095 GTAGGCAGGATGTTGCAGTTAGCCGGC...GGCGTGTCGGATGAGAAGTC 5141
1938 LysGlyArgGlyLeuArgAlaAspGlyArgGlyLeuHisArgGlnLysAl 1954
5142 A...CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAA 5188
1954 aValProValIleProSerAsnArgValValThrAspGlyValGluArgL 1971
5189 CAATCACACCAAAATCGGGGGGCAAGGCTCTATCT..... 5223
1971 euSerValLysMetGlnGlyValGluAlaLeuArgThrGluLeuArgIle 1987
5224 GAGGGAAGTGGTAGGGAAGTCAA 5246
1988 LeuGluAspLeuAspSerAlaValIleGluLysLeuAsnArgArg...Ar 2003
5247 GGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAG. 5295
2003 gAsnArgAspThrAsn.....AspAspGluPheThrArgProA 2016
5296 TGTGAGTGGCTCAGGGCT 5313
2016 laHisGluGlnMetGlnGluValThrThrPheCysSer.....LysAla 2030
5314 GATAATCCAGTGATGGCTCTTGAA..... 5337
2031 AsnSerAlaGlyLeuAlaLeuGluArgAlaValLeuValGluAspAlaIl 2047
5338 CCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGG. 5379
2047 eLysSerGluLysLeuSerLysThrValAsnGluMetValArgLysGlyS 2064
5380 ACCTCTGAAGATGCCGTCGTG.....GAGTACTTG 5409
2064 erThrThrSerGluGluValAlaValAlaLeuSerAspAspGluAlaVal 2080
5410 AAGTATCTGGCTATAGGCATTGAGAGG.....ACATACAG 5444
2081 GluGluIleSerValAlaAspGluArgAspAspSerProLysThrValAr 2097
5445 G.....GCGTTGCTTATGGCTAGAA 5464
2097 gIleSerGluTyrLeuAsnArgLeuAsnSerSerPheGluPheProLysP 2114
5465 ATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTT 5514
2114 roIleValValAspAspAsnLysAspThrGlyGlyLeuThrAsnAlaVal 2130
5515 TATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCAT 5564
2131 ArgGlu.....PheTyrTyrMetGlnGluLeuAlaLeuPheGl 2143
5565 TTTTCATTCAACA.....CAAGACGGCTTGCGTGTG..... 5595
2143 uIleHisSerLysLeuCysThrTyrTyrAspGlnLeuArgIleValAsnP 2160
5596 AGAGACCTACCGTACGTACTCATA 5619
2160 heAspArgSerValAlaProCysSerGluAspAlaGlnLeuTyrValArg 2176
5620 GCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCG...GTGGT 5666
2177 LysAsnGlySerThrIleValGlnGlyLysGluValArgLeuHisIleLy 2193
5667 AGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTG..... 5706
2193 sAspPheHisAspHisAspPheLeuPheAspGlyLysIleSerIleAsnL 2210
5707 GTTTTCCACGATGCCATTAATTTGATA 5733
2210 ysArgArgArgGlyGlyAsnValLeuTyrHisAsp.....AsnLeu... 2223
5734 GGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTT..... 5778
2224 ...AlaPheLeuAlaSerAsnLeuPheLeuAlaGlyTyrProPheSerAr 2239
5779 AAGTCCTTCGAATATAAGTGCTATAATGCTC 5809
2239 gSerPheValPheThrAsnSerSerValAspIleLeuLeuTyrGluAlaP 2256
5810 CCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTT..... 5853
2256 roProGlyGlyGlyLysThrThrThrLeuIleAspSerPheLeuLysVal 2272
5854 AAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAG 5897
2273 PheLysLysGlyGluValSerThrMetIleLeuThrAlaAsnLysSerSe 2289
5898 TTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAA. 5946
2289 rGlnValGluIleLeuLysLysValGluLysGluValSerAsnIleGluC 2306
5947 GGTCTCAACAGTGCTACCACAGTT 5970
2306 ysGlnLysArgLysAspLysArgSerProLysLysSerIleTyrThrIle 2322
5971 AACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTT 6020
2323 AspAlaTyrLeuMetHisHis.....ArgGlyCysAspAlaAspValLe 2337
6021 G...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCG 6067
2337 uPheIleAspGluCysPheMetValHisAlaGlySerValLeuAlaCysI 2354
6068 TCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAG 6117
2354 leGluPheThrArgCysHisLysValMetIlePheGlyAspSerArgGln 2370
6118 ATACCATTCATAAACAGG.....GAGAAGGTGTTTAGGATGGATTG 6158
2371 IleHisTyrIleGluArgAsnGluLeuAspLysCysLeuTyrGlyAspLe 2387
6159 TGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGT 6208
2387 uAspArgPheVal...AspLeuGlnCysArgValTyrGlyAsnIleSerT 2403
6209 ACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATG..... 6249
2403 yrArgCysProTrpAspValCysAlaTrpLeuSerThrValTyrGlyAsn 2419
6250 ACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGT 6290
2420 LeuIleAlaThrValLysGlyGluSerGlu..... 2429
6291 TAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTG 6340
2430 GlyLysSer.....SerMetArgIleAsnGluIleA 2440
6341 GAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATG 6390
2440 snSerValAspAspLeuValProAspValGlySerThrPheLeuCysMet 2456

```
    6391 ACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGA.. 6438
         |||   ||||||   ::::::::::::   :::      ||||||
    2457 LeuGlnSerGluLysLeuGluIleSerLysHisPheIleArgLysGlyLe 2473

    6439 .AAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCA 6487
           :::      |||:::||||||||||||||||||||:::|||:::::
    2473 uThrLysLeuAsnValLeuThrValHisGluAlaGlnGlyGluThrTyrA 2490

    6488 GTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACT 6537
         ::   |||   |||   |||   |||      :::|||:::   |||
    2490 laArgValAsnLeuValArgLeuLysPheGlnGluAspGluProPheLys 2506

    6538 AAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGT 6587
         :::      ||||||   |||:::|||||||||||||||   ||||||::
    2507 SerIleArgHisIleThrValAlaLeuSerArgHisThrAspSerLeuTh 2523

    6588 TTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTA 6637
         :|||   :::|||:::::::::      |||   :::         |||:
    2523 rTyrAsnValLeuAlaAlaArgArgGlyAspAlaThrCysAspAlaIleG 2540

    6638 GCGACGCGTCA 6648
         ::   |||:::
    2540 lnLysAlaAla 2543
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y91969

seq_documentation_block:

```
ID   Y91969 standard; Protein; 309 AA.
XX
AC   Y91969;
XX
DT   19-JUL-2000  (first entry)
XX
DE   PMWaV-2 helicase.
XX
KW   Pineapple mealybug wilt virus 2; PMWaV-2; ORF 1a; helicase;
KW   transgenic pineapple; resistance; antiviral.
XX
OS   Pineapple mealybug wilt virus 2.
XX
PN   WO200017372-A2.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22152.
XX
PR   23-SEP-1998;   98US-0101461.
XX
PA   (UYHA-) UNIV HAWAII.
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Hu JS,  Karasev AV,  Dawson WO,  Melzer M;
XX
DR   WPI; 2000-283596/24.
DR   N-PSDB; A08690.
XX
PT   Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT   useful for protecting pineapples against the virus
XX
PS   Claim 17; Page 97-98; 112pp; English.
XX
CC   Pineapple mealybug wilt virus 2 (PMWaV-2), open reading frame (ORF) 1a
CC   is incomplete but encodes this protein which encompasses all motifs
CC   characteristic of viral helicases.  The DNA and protein sequences
CC   are useful for production of transgenic pineapple plant cultivars and
CC   also in transformation methods to impart resistance against the virus to
CC   pineapple plants.  Mealybug wilt is a major problem limiting profitable
CC   pineapple production in many pineapple growing areas worldwide, the
CC   present invention aims to overcome this problem.
XX
SQ   Sequence   309 AA;
```

```
alignment_scores:
            Quality:  631.50              Length:    304
              Ratio:  2.782                 Gaps:      4
 Percent Similarity:  74.671   Percent Identity: 44.737

alignment_block:
 US-09-301-906-3 x Y91969   ..

 Align seg 1/1  to: Y91969  from: 1  to: 309

    5845 GAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGG 5894
         |||:::   |||:::   ::::::::::::   :::|||||||||   ||
       6 GluTyrPheLysAspArgSerGlySerCysIleMetThrAlaAsnArgGl 22

    5895 AAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCG..... 5937
         ||||:::   |||||||||   ::::::::::::   |||
      22 ySerAlaIleAspIleAsnAspThrIleGluSerIleAspAlaAlaAsnA 39

    5938 ...................AATTTGGAAGGTCTCAACAGT......... 5958
                             |||::::::|||::::::|||
      39 laSerLysAlaAlaSerAsnAsnValSerGlyValGluSerIleAspAsn 55

    5959 .........GCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAG 5999
                  |||   ||||||||||||:::::::::|||         ::
      56 TyrValCysAlaArgThrValAsnSerGlnIleMetAsn......CysLy 70

    6000 GGGAATG......TATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGC 6043
         :|||:::      |||      :::||||||||||||:::::::::||||
      70 sGlyValMetAsnTyrThrCysAlaLeuValAspGluMetTyrLeuMetH 87

    6044 ATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGC 6093
         ||:::|||||||||:::||||||||||||||:::::|||||   :::::
      87 isLysGlyLeuLeuMetLeuGlyValPheSerSerGlyAlaArgArgAla 103

    6094 CTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGT 6143
         :::|||:::||||||||||||||||||||||||||||||||||||||||
     104 IlePheTyrGlyAspIleAsnGlnIleProPheIleAsnArgGluLysCy 120

    6144 GTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTAT 6193
          |||          :::|||:::   |||   |||:::::::::::|
     120 sPheTyrSerLysGluGlyValTyrCysProGlyLysAspGluIleIleT 137

    6194 ACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
         ||||||||:::|||||||||||||||   |||||||||         |||
     137 yrThrSerGluSerTyrArgCysProAlaAspValCysMetTrpSerSer 153

    6244 TCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAG 6293
         |||:::   :::::::::   ::::::   |||   :::   |||
     154 SerLeuLysAlaGlnAlaGlySerAsnArgTyrLeuLysGlyValSerCy 170

    6294 CGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAA 6343
          ::::::   :::   |||:::|||||||||||||||||||||:::
     170 sAsnGlnArgGluValValLeuArgSerLeuSerLysArgProValValT 187

    6344 CCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGACC 6393
           ::::::::::|||   :::::::::||||||:::|||:::   :::
     187 yrAlaGluGlnValIleGlnLeuGluAlaAspAlaTyrIleThrPheLys 203

    6394 CAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGA 6443
         |||      |||:::   :::   |||:::|||::::::   |||:::::
     204 GlnGluCysLysGluLysValValArgAlaLeuArgAlaValGlyArgAr 220

    6444 AACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATG 6493
         :      |||:::|||   ||||||||||||||   ||||||:::   |
     220 gAspLysValPheThrSerHisGluAlaGlnGlyMetThrPheGlyArgV 237

    6494 TGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAA 6543
         ||||||||   |||   :::   :::||||||||||:::|||:::::::::
     237 alValLeuCysArgLeuSerAlaThrAspAspSerValPheSerSerGlu 253

    6544 CCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGC 6593
         ||||||||||||||:::||||||||||||||:::|||   ||||||||
```

```
     254 ProHisIleLeuValAlaLeuSerArgHisThrGlnSerCysValTyrAl 270

    6594 CGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACG 6643
         |:::|||||||||||||||   ||||||||||||:::   |||:::::::
     270 aThrLeuSerSerLysLeuAlaAspLysValGlyAlaAlaIleAspSerV 287

    6644 CGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGCT 6693
         :::::   ::::::|||||||||:::::|||   |||||||:::   :::
     287 alThrArgLysGluValSerAspThrValLeuLysThrPheValAlaSer 303

    6694 GGTTGCTTTCGA 6705
         :::   ||||||
     304 AlaLeuPheArg 307

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y58143

seq_documentation_block:
ID   Y58143 standard; Protein; 120 AA.
XX
AC   Y58143;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein proteinase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49203.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 4; Fig 6; 84pp; English.
XX
CC   This sequence represents the proteinase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   120 AA;


alignment_scores:
            Quality:  605.00              Length:   120
              Ratio:  5.042                Gaps:     0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-301-906-3 x Y58143   ..

 Align seg 1/1  to: Y58143  from: 1  to: 120

     253 GTCAGCGGCTCAGTCAGCGCGCTGAGAGGGGATGGTAAGAAGGTCTTGAT 302
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ValSerGlySerValSerAlaLeuArgGlyAspGlyLysLysValLeuMe 17

     303 GGAGGCAAGGACCTCAACTTCCGCAACTTCCGACGTGTCTGATTTCGACG 352
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tGluAlaArgThrSerThrSerAlaThrSerAspValSerAspPheAspV 34

     353 TCGTATTCGAAGCTGTTTCTAATGCATTACTTGTCGTACACTACCACCGG 402
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 alValPheGluAlaValSerAsnAlaLeuLeuValValHisTyrHisArg 50

     403 GTAGTGCCGTATGCCCCCGTCAAGCGCGAGCAGCCTAAACCGGCTGTTAA 452
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ValValProTyrAlaProValLysArgGluGlnProLysProAlaValLy 67

     453 GCAAGATGAGCAGAAGCCCAAACGGCAAGCGTCACATTGGGCTGTTAAGC 502
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 sGlnAspGluGlnLysProLysArgGlnAlaSerHisTrpAlaValLysP 84

     503 CAACAGCTGTTGGCGTCCACGTACCACTTCCTAAAAAACAGGAAGCACTG 552
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 roThrAlaValGlyValHisValProLeuProLysLysGlnGluAlaLeu 100

     553 GAGCCAGCGCAATCAGTCCCACAACAGTCGTTGGAGGAGAAGGCCGCCTT 602
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GluProAlaGlnSerValProGlnGlnSerLeuGluGluLysAlaAlaLe 117

     603 GACGTTTGGC 612
         ||||||||||
     117 uThrPheGly 120

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:Y91963

seq_documentation_block:
ID   Y91963 standard; Protein; 598 AA.
XX
AC   Y91963;
XX
DT   19-JUL-2000  (first entry)
XX
DE   PMWaV-1 helicase.
XX
KW   Pineapple mealybug wilt virus 1; PMWaV-1; ORF 1a; helicase;
KW   transgenic pineapple; resistance; antiviral.
XX
OS   Pineapple mealybug wilt virus 1.
XX
PN   WO200017372-A2.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22152.
XX
PR   23-SEP-1998;   98US-0101461.
```

```
XX
PA   (UYHA-) UNIV HAWAII.
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Hu JS,  Karasev AV,  Dawson WO,  Melzer M;
XX
DR   WPI; 2000-283596/24.
DR   N-PSDB; A08684.
XX
PT   Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT   useful for protecting pineapples against the virus
XX
PS   Claim 5; Page 79-82; 112pp; English.
XX
CC   Pineapple mealybug wilt virus 1 (PMWaV-1), open reading frame (ORF) 1a
CC   encodes this protein with all 8 motifs conserved in the so-called viral
CC   helicases.  The DNA sequences and related proteins are useful for
CC   production of transgenic pineapple plant cultivars and also in
CC   transformation methods to impart resistance against the virus to
CC   pineapple plants.  Mealybug wilt is a major problem limiting profitable
CC   pineapple production in many pineapple growing areas worldwide, the
CC   present invention aims to overcome this problem.
XX
SQ   Sequence   598 AA;


alignment_scores:
            Quality:  323.00              Length:    490
              Ratio:  1.192                 Gaps:     22
 Percent Similarity:  55.306   Percent Identity: 26.735

alignment_block:
 US-09-301-906-3 x Y91963   ..

 Align seg 1/1  to: Y91963  from: 1  to: 598

    5359 TTTGAAGTGGTTAAAACC..................GGGACCTCTGAAGA 5390
         |||||||||:::||||||                  |||   |||
     117 PheGluValIleLysThrIleLeuAlaAsnThrLeuGlyPheSerPheTh 133

    5391 TGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACAT 5440
          ::::::   ::::::|||                          ::::::|
     133 rGlyMetAsnSerPheLeuAspIle..................GluSerT 144

    5441 ACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGT 5490
         |||||:::|||:::                  |||:::   ::::::|||
     144 yrArgThrLeuValArg............ArgValSerGlySerAspGly 156

    5491 GTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCAC.. 5538
                 |||   :::   :::   :::||||||   |||::::::
     157 TyrSerAlaValLeuGluAlaLeuCysHisSerLeuHisGlyLeuArgAl 173

    5539 .....................GTTTACAAGTCGGGCACAGATCTCATTT 5566
                              |||   |||:::         ||||||
     173 aGluValAspValLeuAspLysValValLysGlnGluIleProLeuIleG 190

    5567 TTCAT.....................TCAACACAAGACGGCTTGCGTGTG 5595
           |||                    ::::::|||:::      ||||||
     190 lyHisLysHisGluLeuPheCysLysAsnValGlnGluLeuAspArgVal 206

    5596 AGAGACCTACCG..................................... 5607
         :::      |||
     207 LysTyrLysProAspGlyPheHisTyrTyrAsnValGluLeuSerAsnIl 223

    5608 .TACGTACTCATA...GCTGAAAAAGGTATCTTTACCAAGGGCAAAGATG 5653
          |||   |||:::   ::::::|||   ::::::::::::
      223 eTyrGlyLeuValAsnThrHisLysLeuValTyrAsnAsn.......... 236

    5654 TCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATA 5703
           |||   ||||||      ::::::::::::   :::   :::
     237 ..AspProValValLysGluSerGluGlyIleValLeuLeuGluPro... 251

    5704 CTGGTTTTCCACGAT...GCCATTAATTTGATAGGTGCACTG........ 5742

            |||:::   ::::::|||||||||   :::|||
     252 .........HisGluIleSerPheAsnLeuIleArgSerLeuAlaLeuIl 265

    5743 .............AAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTA 5779
                      ||||||:::           |||   |||      |
     265 eAspLeuLeuValLysValSer............ValGluGluValAsnL 278

    5780 AGTCCTTCGAA...TATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAG 5826
         ||::::::::::      |||      ||||||   ||||||:::||||||
     278 ysAlaIleAspAsnValLysPheValAsnAlaValProGlyAlaGlyLys 294

    5827 ACGACGACGTTAGTGGACGAATTCGTT..................AAGTC 5858
         |||      :::   ::::::::::::                  |||::
     295 ThrTyrGlnIleLysGlnArgMetLeuArgTrpPheAspSerGluLysAs 311

    5859 ACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACA 5908
         :   ::::::      ::::::|||::::::   :::|||::::::   :
     311 pGlySerAlaLeuLeuValLeuThrSerSerArgAsnSerAlaAspThrL 328

    5909 TAAAT...ATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAAC 5955
         :::::   :::|||   :::||||||         |||      :::
     328 euLysAlaPheAlaGlnGluLysArg.........LeuGlyLysMetIle 341

    5956 AGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGA.. 6003
         :::      ||||||::::::   :::      |||   :::||||||
     342 GlnIleLeuThrValAspAlaPheLeu......PheGlnAlaArgGlyAr 356

    6004 ..........ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGC 6043
                    :::||||||   :::|||:::||||||   :::|||   |
     356 gAsnValArgLeuTyrLysThrLeuLeuIleAspGluCysTyrMetThrH 373

    6044 ATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGC 6093
         ||   |||:::|||:::      :::   |||:::      :::|||
     373 isAlaGlyIleLeuArgGlyIleIleAlaAlaValLysProGluGluCys 389

    6094 CTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGT 6143
         :::::::::||||||   :::|||:::||||||||||||||||   |||::
     390 ValLeuTyrGlyAspArgArgGlnValProPheIleAsnArgIleLysLe 406

    6144 GTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTAT 6193
         :::::::   :::   :::   :::   |||:::      :::
     406 uLeuAsnAspAsnLysSerPheLeuLysProSerLeuGlyAsnTyrSerG 423

    6194 ACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
                  :::   |||||||||   |||:::|||:::   :::|||
     423 luMetLeuIleThrArgArgCysProAlaAspIleCysTrpArgMetSer 439

    6244 TCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAG 6293
         :::      |||::::::   :::|||         :::::::::   ||
     440 Asn......ValAsnAsnGlyLysLys......GlyAspArgLeuTyrSe 452

    6294 CGGT..................AAGGACAAACCAGTAGTAAGATCGCTGT 6325
         ||||                 ::::::||||||||:::::::|||::::
     452 rGlyProValLysLeuPheThrGlnSerLysProValLeuLysSerValT 469

    6326 CCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGAC 6375
         ::   :::   ::::::      |||                 :::::|||
     469 hrCysLysAlaPheSerLysGlyAspHisAsnLeuPheSerGlnValAsp 485

    6376 GTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTT 6425
               ::: . :::||||||   ||||||::::::::      :::
     486 ArgValMetThrPheThrGlnAsnGluLysAsnGluLeuIleSerGluTy 502

    6426 GAAGGGAAAAGGA..................AAAGAAACACCAGTGATGA 6457
              ::::::|||                  :::|||   :::   |
     502 rMetSerArgGlyIleGlyThrIleGlnAspAlaLysThrLeuIleGlyT 519

    6458 CAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGG 6507
         |||||   |||:::||||||:::|||::::::   |||   |||
     519 hrValAlaGluSerGlnGlyGluThrTyrLysArgValHisLeuValAla 535

    6508 ACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGT 6557
            |||   :::||||||:::::|||:::::::::||||||   |||||
```

```
     536 PheLysProThrAspAspGlnValPheSerSerMetProHisArgLeuVa 552

    6558 TGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAA 6607
         |:::|||||||||||||||   ||||||   |||   ::::::   :::|
     552 lAlaLeuSerArgHisThrIleSerLeuGlnTyrPheCysIleProAsnL 569

    6608 AGTTGGACGATAAGGTCGGC 6627
         ||::::::      :::|||
     569 ysMetAsnLysGlyIleGly 575

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R05107

seq_documentation_block:
ID   R05107 standard; protein; 1128 AA.
XX
AC   R05107;
XX
DT   05-OCT-1990  (first entry)
XX
DE   Sequence encoded by RNA1 of AMV.
XX
KW   Alfalfa mosaic virus; AMV; cryptic infection; probe.
XX
OS   Nicotiniana tabacum L "Xanthi-nc".
XX
PN   US4921802-A.
XX
PD   01-MAY-1990.
XX
PF   13-JAN-1988;   88US-0144692.
XX
PR   05-MAY-1983;   83US-0492582.
PR   13-JAN-1988;   88US-0144692.
XX
PA   (PIOE ) PIONEER HI-BRED INTERNATIONAL LTD.
XX
PI   Hall TC,  Loesch-Fries SL,  Jarvis NP,  Barker RF;
XX
DR   WPI; 1990-171297/22.
DR   N-PSDB; Q04476.
XX
PT   Recombinant cDNA plasmids derived from alfalfa mosaic virus RNA -
PT   for detection of cryptic infection by alfalfa mosaic virus.
XX
PS   Disclosure; ; p; English.
XX
CC   The sequence is encoded by the bottom (B)- or "RNA1" genome fragment
CC   of AMV which is necessary to initiate infection.   DNA encoding the
CC   sequence can be used as a probe to detect cryptic infection of plants
CC   by AMV.
CC   See also R05108-09.
XX
SQ   Sequence   1128 AA;


alignment_scores:
           Quality:  323.00            Length:   1871
             Ratio:  0.490               Gaps:     65
 Percent Similarity:  35.222   Percent Identity: 17.050

alignment_block:
 US-09-301-906-3 x R05107   ..

 Align seg 1/1  to: R05107  from: 1  to: 1128

    1303 ACGAAATTGTGCGATATGTTTTCTCAA.................AGGGA 1334
         |||   :::      :::||||||:::               :::||
      40 ThrAlaIleGlyLysValPheSerGluAlaGlyArgAlaTyrAlaGlnAs 56

    1335 CGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTT..... 1377
         ||||:::           ||||||   :::   :::|||:::
      56 pAlaLeu...........ProSerAspLysGlyGluValLeuLysIleS 69

    1378 ........CTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCA 1419
                 |||         :::::::::   ||||||      ||||||
      69 erPheSerLeuAspAlaThrGlnGlnAsnIleLeuArgAlaAsnPhePro 85

    1420 GAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGC 1469
                |||   ||||||:::|||      ||||||||   |||||
      86 GlyArgArgIleValPheSerAsnSerSerSerSerSerHisCysPheAl 102

    1470 TAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTA 1519
         |   |||   |||         :::   :::      |||      ||||
     102 aAlaAlaHisArgLeuLeuGluThrAspPheValTyrArgCysPheGlyA 119

    1520 ATGTG......TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTC 1563
         ||:::         :::::|||:::||||||:::|||:::   |||:::
     119 snThrValAspSerIleIleAspLeuGlyGlyAsnPheValSerHisMet 135

    1564 AAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAA 1613
         |||:::   :::   |||   |||   |||   |||:::||||||:::::
     136 LysValLysArgHisAsnValHisCysCysCysProIleLeuAspAlaAr 152

    1614 AGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGG 1663
         :|||      |||      ::::::||||||   |||:::
     152 gAspGlyAlaArgLeuThrGluArgIleLeuSerLeuLys..... 165

    1664 GAGATTCGTACGTGTCCAGT......GACCTTCTAACTGAAGCGGCTTCA 1707
              |||||||||   :::      :::::::::   ||||||
     166 .....SerTyrValArgLysHisProGluIleValGlyGluAla..... 178

    1708 AAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGC 1757
                 :::||||||            |||:::|||:::   |||||
     179 .........AspTyrCysMetAspThrPheGlnLysCysSerArgArgAl 192

    1758 CGATGCGGGTTTTATGGTG...GATGTGTACGATATATCCCCGCAGCAGG 1804
         ||||   :::|||   :::      :::   |||::::::      ::::
     192 aAspTyrAlaPheAlaIleHisProThrSerAspLeuAspValGlyGluL 209

    1805 TAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTT 1854
         ::|||   ::::::|||:::||||||:::::   |||      ::::::
     209 euAlaCysSerLeuAspGlnLysGlyValMetLysPheIleCysThrMet 225

    1855 ATGTTCCCCGTGGAG.........TTGTTGTACGGTAACGGTGAAGTTTA 1895
         |||      ::::::         |||:::::::::::::||||||:::
     226 MetValAspAlaAspAlaAspMetLeuIleHisSerGluGlyGluIlePr 242

    1896 CTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCT 1945
              ::::::         :::   ::::::::::|||   :::   :
     242 oAsnPheAsnValArgTrpGluIleAspHisLysAsnAspLeuIleHisP 259

    1946 ACAAT...GTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAAC 1992
         :::::   :::   :::   ::::::   |||:::|||   ||||||
     259 heAspPheIleAspGluProAsnSerGlyTyrSerHisArgPheSerLeu 275

    1993 GTAAGCGGGTTTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGT 2042
         ::::::   ::::::|||:::                        |||::
     276 LeuLysHisTyrLeuThrTyr.......................AsnAl 284

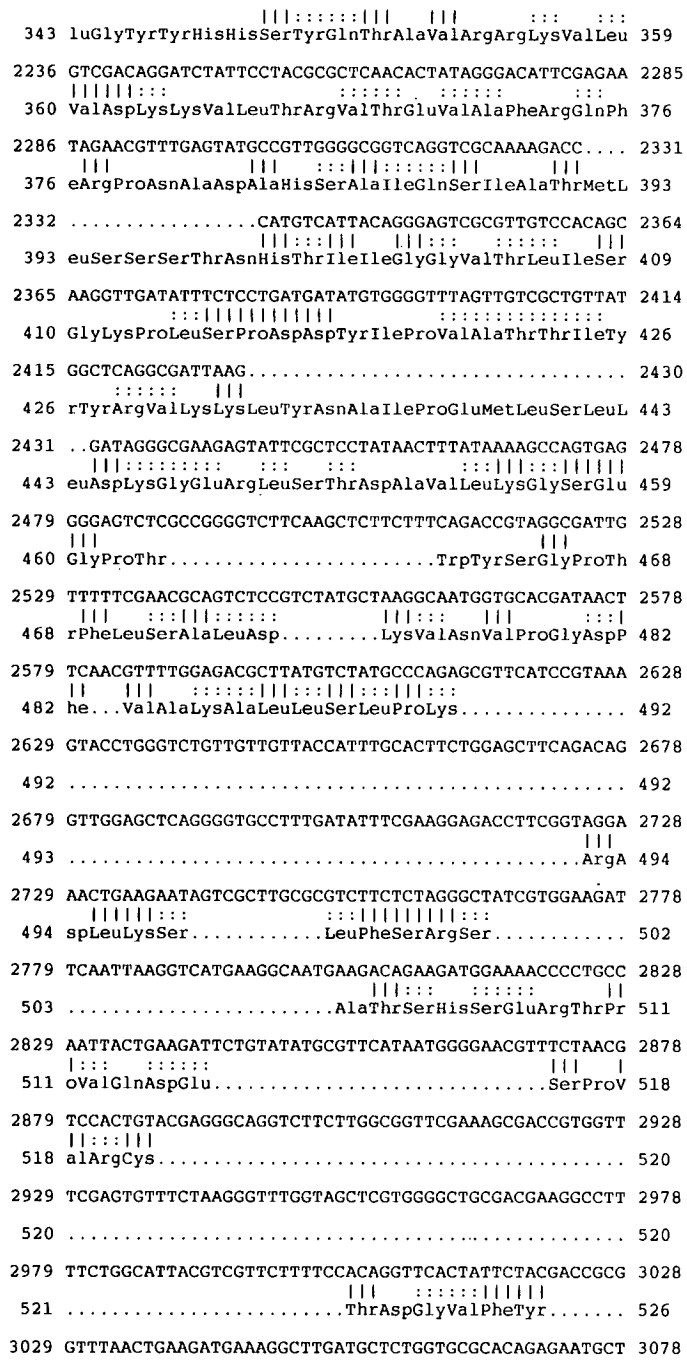
    2043 GTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTA 2092
         :      |||   :::   :::||||||                    :
     284 aValAspLeuGlyHisAlaAlaTyrArg...................I 294

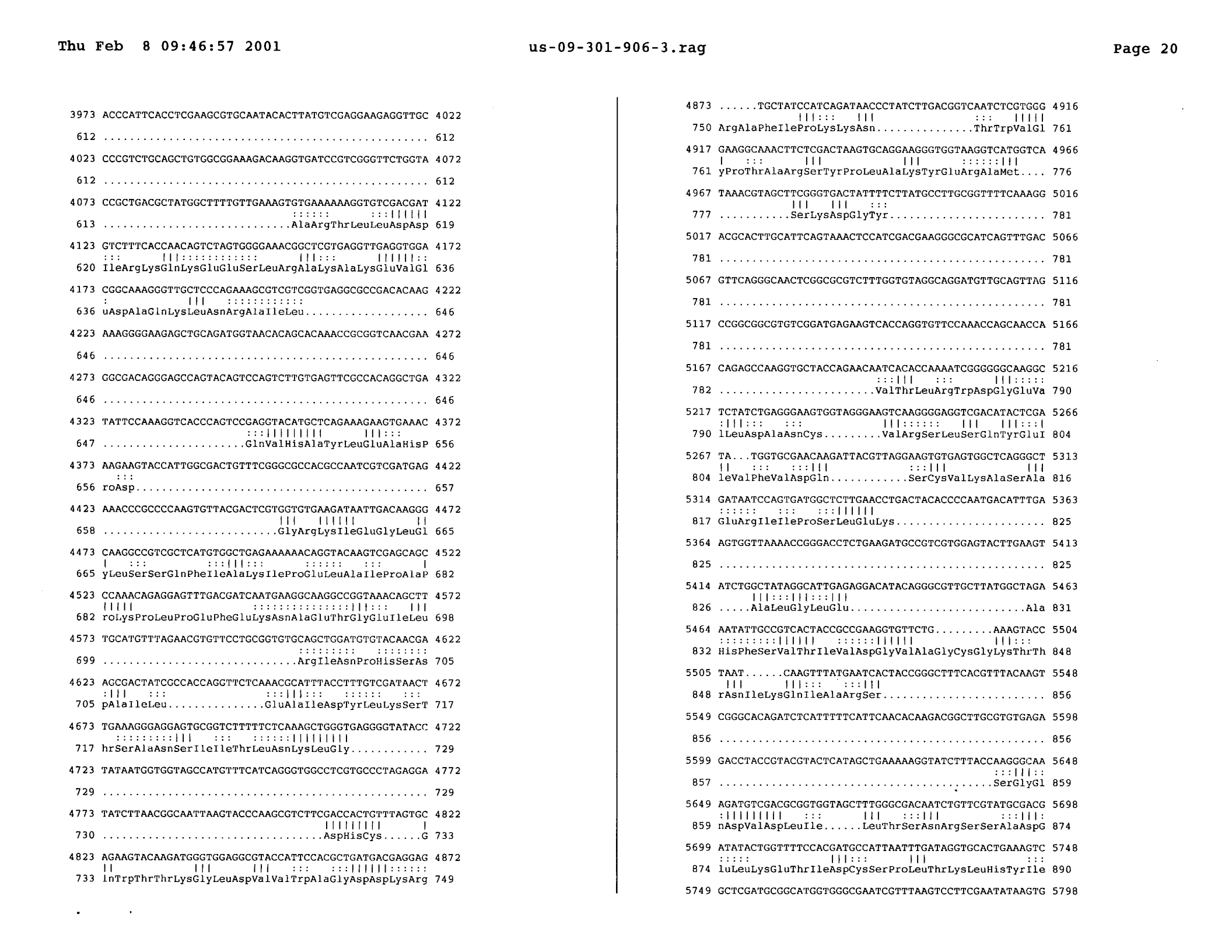
    2093 TGTGTAGGGCTCAGAAGTCACCTGGAACT.........GAGGTTACGTAT 2133
         ::   |||   |||         |||:::         ::::::||||||
     294 leGluArgLysGlnAspPheGlyGlyValMetValIleAspLeuThrTyr 310

    2134 AGGTCGTTG..............GTCCCGTCGTTCGTGGGCAAATCGCT 2168
            ||||||               :::|||         |||:::|||
     311 ...SerLeuGlyPheValProLysMetProHisSerAsnGlyArgSerCy 326

    2169 GGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG.............. 2202
          ::::::         |||   |||:::   |||
     326 sAlaTrpTyrAsnArgValLysGlyGlnMetValValHisThrValAsnG 343

    2203 ................TCCTTTAAGACAATAGTCCTCGATTCGGACTTT 2235
```


343 luGlyTyrTyrHisHisSerTyrGlnThrAlaValArgArgLysValLeu 359
2236 GTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGACATTCGAGAA 2285
360 ValAspLysLysValLeuThrArgValThrGluValAlaPheArgGlnPh 376
2286 TAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACC.... 2331
376 eArgProAsnAlaAspAlaHisSerAlaIleGlnSerIleAlaThrMetL 393
2332 CATGTCATTACAGGGAGTCGCGTTGTCCACAGC 2364
393 euSerSerSerThrAsnHisThrIleIleGlyGlyValThrLeuIleSer 409
2365 AAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTAT 2414
410 GlyLysProLeuSerProAspAspTyrIleProValAlaThrThrIleTy 426
2415 GGCTCAGGCGATTAAG..... 2430
426 rTyrArgValLysLysLeuTyrAsnAlaIleProGluMetLeuSerLeuL 443
2431 ..GATAGGGCGAAGAGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAG 2478
443 euAspLysGlyGluArgLeuSerThrAspAlaValLeuLysGlySerGlu 459
2479 GGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGATTG 2528
460 GlyProThr.....TrpTyrSerGlyProTh 468
2529 TTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACT 2578
468 rPheLeuSerAlaLeuAsp.....LysValAsnValProGlyAspP 482
2579 TCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAA 2628
482 he...ValAlaLysAlaLeuLeuSerLeuProLys..... 492
2629 GTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAG 2678
492 492
2679 GTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGA 2728
493 ArgA 494
2729 AACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGAT 2778
494 spLeuLysSer.....LeuPheSerArgSer..... 502
2779 TCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCC 2828
503 AlaThrSerHisSerGluArgThrPr 511
2829 AATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAACG 2878
511 oValGlnAspGlu.....SerProV 518
2879 TCCACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTT 2928
518 alArgCys..... 520
2929 TCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTT 2978
520 520
2979 TTCTGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCG 3028
521 ThrAspGlyValPheTyr..... 526
3029 GTTTAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCT 3078
526 526
3079 ATAAACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAA 3128
527 ProIleArgMetLeu.....LeuLy 533
3129 GGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGACT 3178
533 sCysLeuGlySerAspLys.....PheGluSerValThrIleThrAsp. 547
3179 TCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCG 3228
547 547
3229 TCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGC 3278
547 547
3279 TAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAA 3328
547 547
3329 GTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCC 3378
548 ProArgSerAsnThrGluThrThrValAsp 557
3379 GGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCAC 3428
558 LeuTyrHisSerPheGlnLysLysIle..... 566
3429 TAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTT 3478
566 566
3479 CTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACAT 3528
566 566
3529 GCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGG 3578
567 GluThrValPheSerPheIleLeuGlyLysIleAspGl 579
3579 GGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATT 3628
579 yPro..... 580
3629 GGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTG 3678
580 580
3679 TTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAA 3728
580 580
3729 CGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATG 3778
581 SerProLeuIleSerAspProV 588
3779 TGTACGATGAGACGCTTAGG.....TATTACCGGGACTTTGACTATGAC 3822
588 alTyrPheGlnSerLeuGluAspValTyrTyrAlaGluTrp..... 601
3823 GAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGA 3872
601 601
3873 TAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATG 3922
602 .HisGlnGlyAsnAlaIleAspAlaSerAsnTyr..... 612
3923 TGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAG 3972
612 612



```
          |||
 891 ArgThrCys..................................... 893

5799 CTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAAT 5848
                                                 |||::::
 894 .....................................AspSerT 896

5849 TCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGT 5898
     ::
 896 yr.............................................. 896

5899 TCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGG 5948

 896 ..............................................  896

5949 TCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCA 5998
      |||   ||||||::::::|||::::::
 897 .LeuMetSerAlaSerAlaValLysAla..................... 905

5999 GGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAA 6048
               :::|||::::::   ||||||      :::::::|||
 906 ...........GlnArgLeuIlePheAspGluCysPheLeuGlnHisAla 918

6049 GGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTT 6098
     ||||||:::      ::::::      :::|||:::||||||   :::
 919 GlyLeuValTyrAlaAlaAlaThrLeuAlaGlyCysSerGluValIleGl 935

6099 TTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTA 6148
      |||||||||   :::||||||||||||:::::::|||:::      |||
 935 yPheGlyAspThrGluGlnIleProPheValSerArgAsnProSerPheV 952

6149 GGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACT 6198
       :::          ::::::   |||   |||      :::   |||
 952 alPheArgHisHisLysLeuThrGlyLysValGluArgLysLeuIleThr 968

6199 TCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAAT 6248
           :::|||   |||   |||:::   |||   |||
 969 .........TrpArgSerProAlaAspAlaThrTyrCysLeu........ 979

6249 GACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTA 6298
                                  ||||||           :
 980 ............................GluLysTyrPheTyrLysA 986

6299 AGGACAAACCA.................GTAGTAAGATCGCTGTCCAAA 6330
     ::   ||||||               |||:::||||||::::::
 986 snLysLysProValLysThrAsnSerArgValLeuArgSerIleGluVal 1002

6331 AGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTA 6380
        ||||||::::::      :::|||   :::   :::      :::||
1003 ValProIleAsnSerProValSerValGluArgAsnThrAsnAlaLeuTy 1019

6381 CTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGG 6430
     |||||||   ||||||   ||||||:::   :::|||   :::
1019 rLeuCysHisThrGlnAlaGluLysAlaValLeuLysAlaGlnThrHisL 1036

6431 GAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAA 6480
       ||||||   ::::::   ::::::|||:::||||||||||||||||||
1036 euLysGlyCysAspAsn...IlePheThrThrHisGluAlaGlnGlyLys 1051

6481 ACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCT 6530
     ||||||:::::|||   :::   |||      :::::::::   |||||
1052 ThrPheAspAsnValTyrPheCysArgLeuThrArgThrSerThrSerLe 1068

6531 ATTCACT...AAACAACCG..............CATATACTTGTTGGTT 6562
     |   |||   ::::::|||               |||   ||||||:::|
1068 uThrThrGlyArgAspProIleAsnGlyProCysHisGlyLeuValAlaL 1085

6563 TGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTG 6612
     ||||||||||||   :::::::::   |||   :::::::::
1085 euSerArgHisLysLysThrPheLysTyrPheThrIleAlaHisAspSer 1101

6613 GACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCAGTATC 6662
     ||||||   :::   :::         ||||||:::      ::::::::
```

```
1102 AspAspValIleTyrAsnAlaCysArgAspAlaGly......AsnThrAs 1116

6663 CGACGCTTTGCTT 6675
     :|||:::::::|||
1116 pAspSerIleLeu 1120
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R60608

seq_documentation_block:

ID R60608 standard; Protein; 1116 AA.
XX
AC R60608;
XX
DT 21-JUN-1995 (first entry)
XX
DE Tobamovirus replication enzyme subunit.
XX
KW Tobamovirus; tobacco mosaic virus; vector; transcription; promoter;
KW terminator; replication enzyme; replication site; ori; coat protein.
XX
OS Tobacco mosaic virus.
XX
PN JP06233688-A.
XX
PD 23-AUG-1994.
XX
PF 29-NOV-1993; 93JP-0297704.
XX
PR 01-DEC-1992; 92JP-0322100.
XX
PA (SANY) SANKYO CO LTD.
XX
DR WPI; 1994-319190/40.
DR N-PSDB; Q71546.
XX
PT Improved transcriptional vector of plant virus - esp. vectors
PT pL5013 and pL5020 which are prepd. and mass produced, then
PT tested for stability and function
XX
PS Disclosure; Page 17-19; 36pp; Japanese.
XX
CC A transcriptional vector comprises in a 5'-3' direction (i) a 5'
CC non-transcribed region of tobamovirus, (ii) a replicative enzyme of
CC tobamovirus, (iii) a foreign gene connecting region, and (iv) a 3'
CC non-transcribed region of tobamovirus. The vector also has a
CC transcriptional promoter region upstream of the tobamovirus
CC region and a transcriptional terminator region downstream of the
CC tobamovirus region, neither of which is recognised by the host cell,
CC a replication initiation site which functions in the host cell, and
CC a selective marker gene. This sequence is a subunit of the
CC tobamovirus replication enzyme described (See also R60609) and the
CC sequence encoding it was used in a construct (See Q71546) which in
CC turn was used in the construction of the transcriptional vectors
CC pL5013 and pL5020.
XX
SQ Sequence 1116 AA;

alignment_scores:

```
            Quality:  309.00              Length:   1879
              Ratio:  0.445                 Gaps:     64
 Percent Similarity:  36.935   Percent Identity: 16.764
```

alignment_block:

US-09-301-906-3 x R60608 ..

Align seg 1/1 to: R60608 from: 1 to: 1116

```
1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
     |||   :::|||      :::   |||:::::::::::::::::::   ::
   2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGl 18

1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
```

```
    :::   :::|||   ::::::   |||              ||||||
 18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
     :::   |||   |||:::   ||||||           ::::::|||
  35 laVal...AspGluPheAsnAlaArgAsp.....ArgArgPro 46

1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
     :::   :::   :::      ::::::   ::::::
  47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
     :::   :::||||||::::::|||   |||   ::::::   ::::::
  63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
       |||   :::|||:::::::::|||||      ::::::   ::::::
  80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
       :::         :::         :::|||||||||||||:::|||::
  97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
     :   |||:::      |||::::::      |||   |||   |||   |
 113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.....ATC 1641
     ||||||||:::|||:::   |||:::   ::::::
 130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCC..... 1680
     |||:::||||||            |||:::::::::|||
 147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

1681 .....AGTGACCTTCTAACTGAAGCGGCTTCAA 1708
                  :::::::::   :::|||::::::
 163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

1709 AGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC...GATTCTAGA 1755
       ::::::           |||:::||||||:::|||      :::|||
 180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196

1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
     :::   |||      :::   ::::::||||||||      ::::::
 197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
       :::   |||:::   :::|||::::::   |||         |||:::
 213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
       |||      :::||||||      :::::::::|||   |||:::|||
 230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
     :::::::::      :::|||:::||||||   |||::::::::::   ::
 247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
     ::::   :::   :::   |||:::||||||:::||||||:::      :
 263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280

2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
     ::         :::|||      :::|||:::   :::|||:::
 280 yrValCysLysThrTyrProProAlaSerAsnArgGluValTyrMetLys 296

2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
     |||:::
 297 GluPhe..... 298

2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

 298 ..... 298

2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

 298 ..... 298

2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                    :::|||   |||:::   ::
 299 .....LeuValThrArgValAsnTh 305

2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
     ::::         ::::::   |||   ||||||
 305 rTrpPheCysLysPheSerArgIleAspThrPhe..... 316

2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                         :::
 317 .....LeuLeuTyr 319

2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
     :::   |||:::|||:::   ||||||
 320 LysGlyValAlaHisLysGlyValAsp..... 328

2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446

 328 ..... 328

2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
          |||   :::|||   ||||||   |||   :::
 329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341

2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
               :::   :::|||::::::   |||   |||:::   :::
 342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
       :::       ::::::   |||:::       |||       ::::::   :
 355 uLeuGluAspSerSerSerValAsnTyrTrpProProLysMetArgAspM 372

2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
     :::::   :::|||
 372 etValIleValPro..... 376

2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

 377 .....Le 377

2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
       ||||||||||||   ||||||                |||::::::|||
 377 uPheAspIleSerLeuGluThr.....SerLysArgThrArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
       :::|||:::         :::   :::|||   :::      |||:::
 390 ysGluValLeu.....ValSerLysAspPheValTyrThrValLeu 403

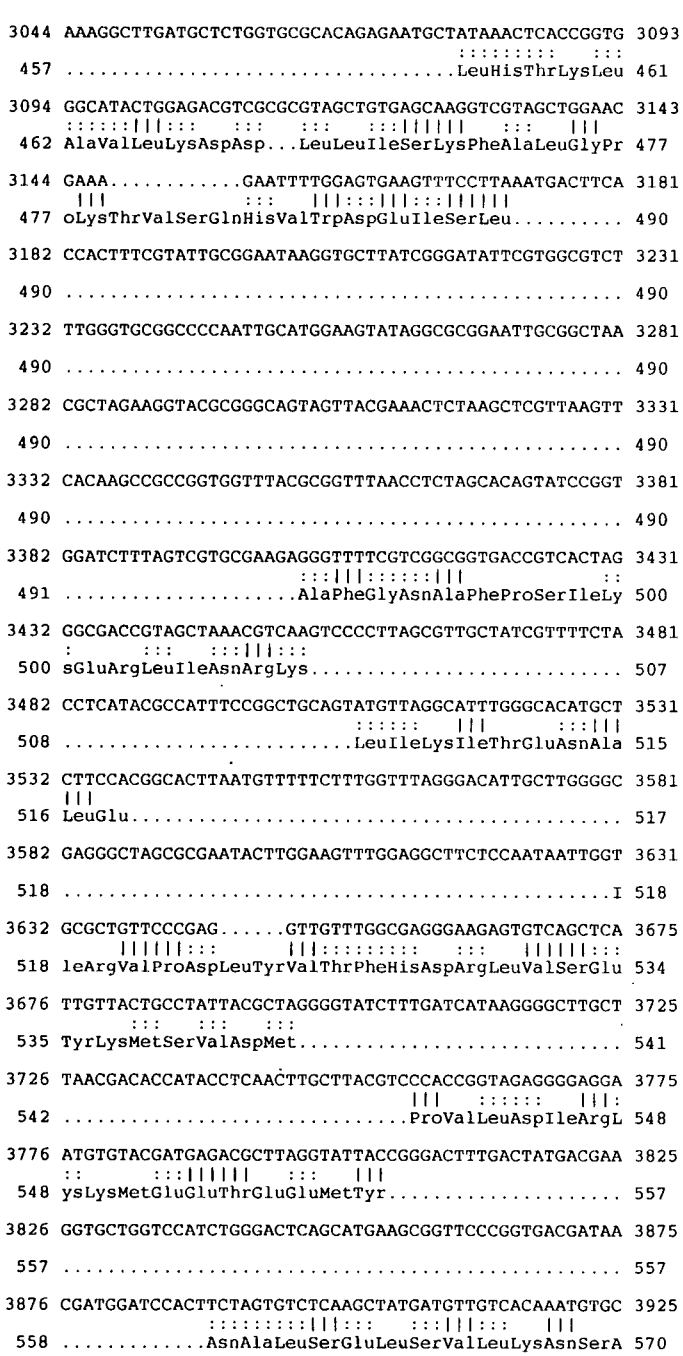
2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   ::::::|||   :::::::|||   |||      |||   :::::
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu.....ThrTyrSerAs 418

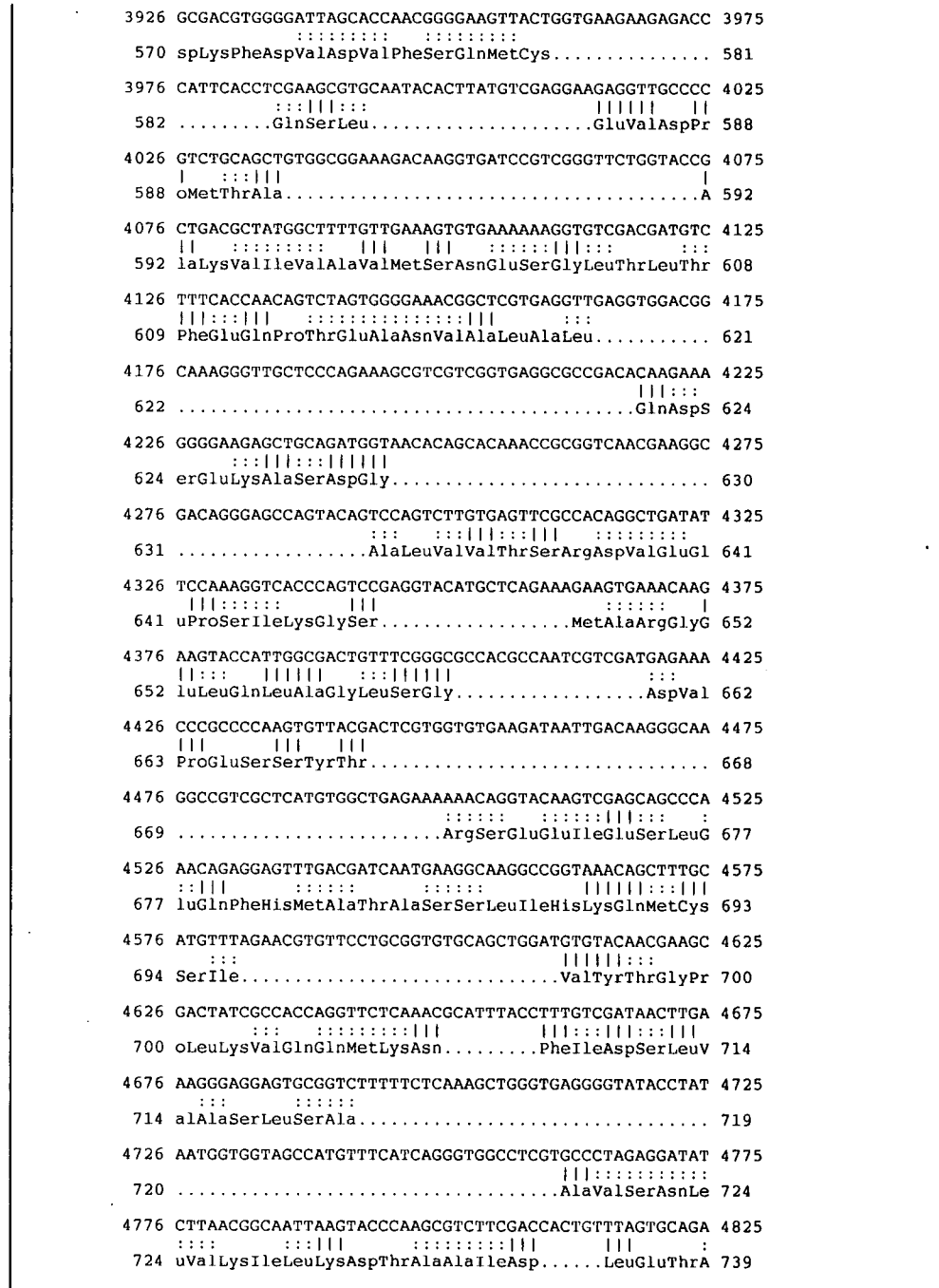
2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::      |||:::|||
 418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430

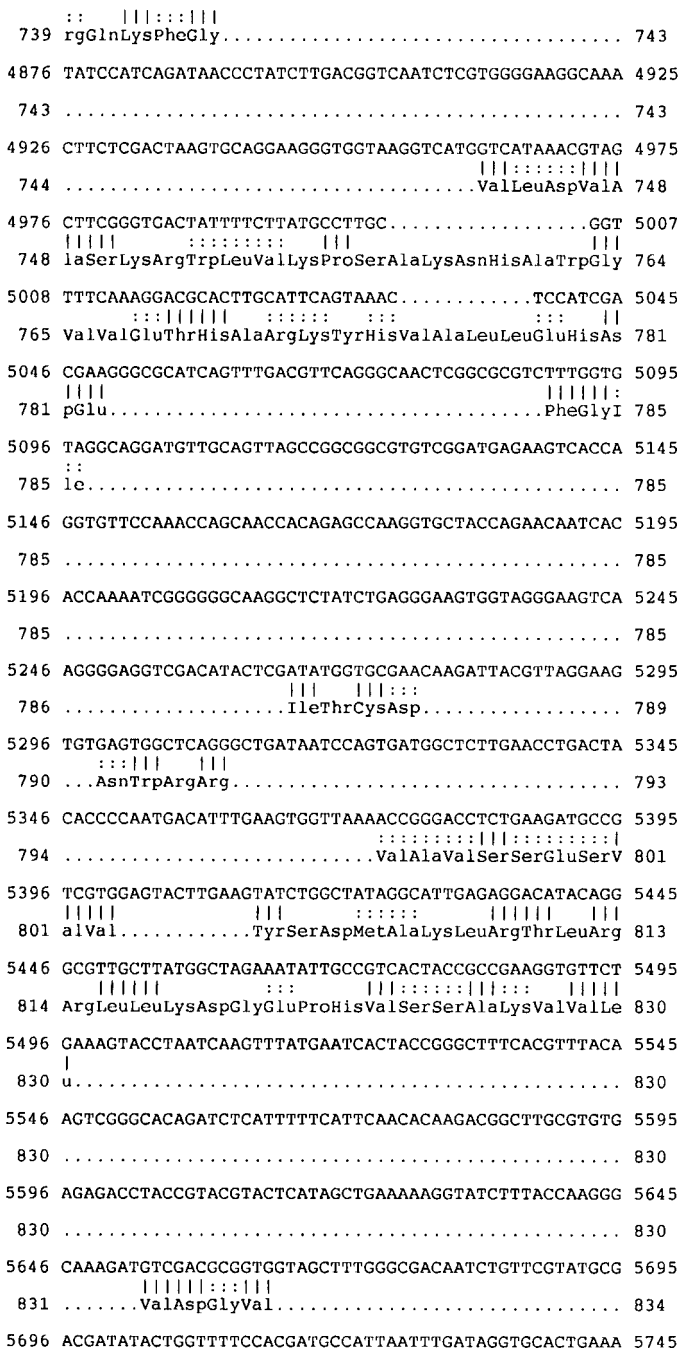
2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
          :::::::::|||      |||              :::|||:::|||
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

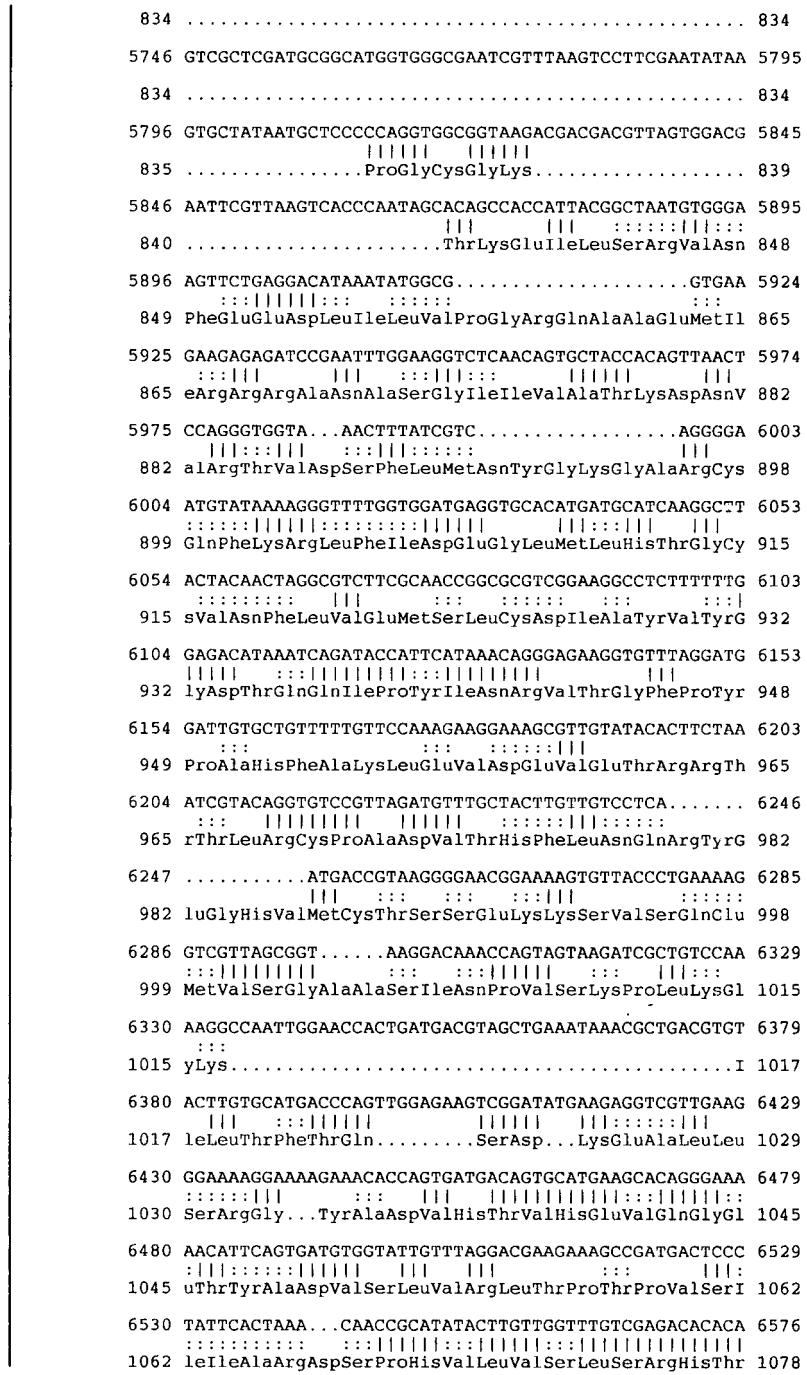
2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||:::         :::   |||
 446 SerLeuLeuGlnSerLeuSerMetThr..... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
      ||||||
 455 .PhePhe..... 456
```


3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
457 LeuHisThrLysLeu 461
3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477
3144 GAAA.....GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu..... 490
3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231
490 490
3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281
490 490
3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331
490 490
3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381
490 490
3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
491 AlaPheGlyAsnAlaProSerIleLy 500
3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
500 sGluArgLeuIleAsnArgLys..... 507
3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
508 LeuIleLysIleThrGluAsnAla 515
3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
516 LeuGlu..... 517
3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631
518 I 518
3632 GCGCTGTTCCCGAG.....GTTGTTTGGCGAGGGAAGAGTGTCAGCTCA 3675
518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534
3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
535 TyrLysMetSerValAspMet..... 541
3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
542 ProValLeuAspIleArgL 548
3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
548 ysLysMetGluGluThrGluGluMetTyr..... 557
3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875
557 557
3876 CGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGC 3925
558 AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570


3926 GCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACC 3975
570 spLysPheAspValAspValPheSerGlnMetCys..... 581
3976 CATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCC 4025
582 GlnSerLeu.....GluValAspPr 588
4026 GTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG 4075
588 oMetThrAla.....A 592
4076 CTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTC 4125
592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuThr 608
4126 TTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGG 4175
609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu..... 621
4176 CAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAA 4225
622 GlnAspS 624
4226 GGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGC 4275
624 erGluLysAlaSerAspGly..... 630
4276 GACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATAT 4325
631 AlaLeuValValThrSerArgAspValGluGl 641
4326 TCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAG 4375
641 uProSerIleLysGlySer.....MetAlaArgGlyG 652
4376 AAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425
652 luLeuGlnLeuAlaGlyLeuSerGly.....AspVal 662
4426 CCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAA 4475
663 ProGluSerSerTyrThr..... 668
4476 GGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCA 4525
669 ArgSerGluGluIleGluSerLeuG 677
4526 AACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGC 4575
677 luGlnPheHisMetAlaThrAlaSerSerLeuIleHisLysGlnMetCys 693
4576 ATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGC 4625
694 SerIle.....ValTyrThrGlyPr 700
4626 GACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGA 4675
700 oLeuLysValGlnGlnMetLysAsn.....PheIleAspSerLeuV 714
4676 AAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTAT 4725
714 alAlaSerLeuSerAla..... 719
4726 AATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATAT 4775
720 AlaValSerAsnLe 724
4776 CTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGA 4825
724 uValLysIleLeuLysAspThrAlaAlaIleAsp.....LeuGluThrA 739
4826 AGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGC 4875


rgGlnLysPheGly
TATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAA
CTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAG
ValLeuAspValA
CTTCGGGTGACTATTTTCTTATGCCTTGC
GGT
laSerLysArgTrpLeuValLysProSerAlaLysAsnHisAlaTrpGly
TTTCAAAGGACGCACTTGCATTCAGTAAAC
TCCATCGA
ValValGluThrHisAlaArgLysTyrHisValAlaLeuLeuGluHisAs
CGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTG
pGlu
PheGlyI
TAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCA
lc
GGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCAC
ACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCA
AGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAG
IleThrCysAsp
TGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTA
...AsnTrpArgArg
CACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCG
ValAlaValSerSerGluSerV
TCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGG
alVal
TyrSerAspMetAlaLysLeuArgThrLeuArg
GCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCT
ArgLeuLeuLysAspGlyGluProHisValSerSerAlaLysValValLe
GAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACA
u
AGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTG
AGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGG
CAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCG
.....ValAspGlyVal
ACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAA


GTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAA
ProGlyCysGlyLys
GTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACG
AATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA
ThrLysGluIleLeuSerArgValAsn
AGTTCTGAGGACATAAATATGGCG
GTGAA
PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetIl
GAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACT
eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV
CCAGGGTGGTA
AACTTTATCGTC
AGGGGA
alArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys
ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTT
GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyCy
ACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTG
sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG
GAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATG
lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr
GATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAA
ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh
ATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA
rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG
ATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAG
luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnClu
GTCGTTAGCGGT
AAGGACAAACCAGTAGTAAGATCGCTGTCCAA
MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl
AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT
yLys
ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG
leLeuThrPheThrGln
SerAsp
LysGluAlaLeuLeu
GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA
SerArgGly
TyrAlaAspValHisThrValHisGluValGlnGlyGl
AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC
uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI
TATTCACTAAA
CAACCGCATATACTTGTTGGTTTGTCGAGACACACA
leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr

```
    6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
         :::||||||   |||   :::              |||   |||   ::
    1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleIl 1095

    6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
         :   |||      :::|||:::::::|||:::   |||
    1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:W87724

seq_documentation_block:
ID   W87724 standard; Protein; 2161 AA.
XX
AC   W87724;
XX
DT   26-APR-1999  (first entry)
XX
DE   Rupestris stem pitting associated virus RSPaV-1 replicase.
XX
KW   RSPaV-1; grape; transgenic plant; disease resistance; replicase.
XX
OS   Rupestris stem pitting associated virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1931
FT                   /note= "encoded by AG, apparent 1 nucleotide
FT                          deletion in open reading frame"
XX
PN   WO9852964-A1.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998;   98WO-US10391.
XX
PR   17-DEC-1997;   97US-0069902.
PR   20-MAY-1997;   97US-0047147.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Meng B;
XX
DR   WPI; 1999-045297/04.
DR   N-PSDB; V99284.
XX
PT   Isolated proteins from Rupestris stem pitting-associated virus and
PT   related nucleic acid - vectors, host cells and transgenic Vitis
PT   cultivars that are resistant to the virus
XX
PS   Claim 4; Page 18-25; 163pp; English.
XX
CC   This is the amino acid sequence of the replicase of Rupestris stem
CC   pitting associated virus RSPaV-1.  It was deduced from the
CC   nucleotide sequence of an open reading frame (ORF1) identified
CC   in the RSPaV-1 genome (see V99284).  Proteins and polypeptides
CC   of RSPaV (see W87724-38), and nucleic acids encoding them (see
CC   V99284-86) are claimed, as are expression systems, transformed
CC   host cells, transgenic grape (Vitis) scion or rootstock cultivars
CC   comprising DNA encoding a replicase, coat protein, or a protein
CC   of a triple gene block, and a method of imparting RSPaV resistance
CC   to Vitis scion or rootstock cultivars by transformation with such
CC   DNA.
XX
SQ   Sequence   2161 AA;


alignment_scores:
           Quality:  290.00            Length:    142
             Ratio:  2.736               Gaps:      3
 Percent Similarity:  74.648   Percent Identity: 41.549

alignment_block:
 US-09-301-906-3 x W87724   ..
```

```
 Align seg 1/1  to: W87724  from: 1  to: 2161

    4660 TTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGG 4709
         |||   |||:::|||:::||||||:::|||   ||||||||||||   ||
     733 PheLysAspArgLeuArgGlyArgGlyAlaSerPhePheSerLysProGl 749

    4710 T...GAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGC 4756
         |   :::   ::::::||||||||||||||||||||:::|||   |||||||
     749 yIleSerCysHisSerTyrAsnGlyGlySerHisThrSerLeuGlyTrpP 766

    4757 CTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTC 4806
         ||:::   ::::::::::||||||:::::           :::   :::
     766 roLysPheMetAspGlnIleLeuSerSerThrGlyGlyArgAsnTyrTyr 782

    4807 GACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCA 4856
         :::   ||||||:::|||   |||:::   ::::::   :::   :::||
     783 AsnSerCysLeuAlaGlnIleTyrGluGluAsnSerLysLeuAlaLeuHi 799

    4857 CGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCA 4906
         |   |||||||||:::||||||         :::   :::|||||||||||
     799 sLysAspAspGluSerCysTyrGluIleGlyHisLysValLeuThrValA 816

    4907 ATCTCGTGGGGAAGGCAAACTTCTCG...ACTAAGTGCAGGAAG..... 4947
         |||||:::|||:::|||:::|||:::   :::|||   |||:::
     816 snLeuIleGlySerAlaThrPheThrIleSerLysSerArgAsnLeuVal 832

    4948 GGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTAT 4997
         ||||||:::         ::::::::::::   :::::::::|||   ||
     833 GlyGlyAsnHisCysSerLeuThrIleGlyProAsnGluPhePheGluMe 849

    4998 GCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACG 5047
         ||||   |||:::|||   ::::::::::|||:::|||::::::
     849 tProArgGlyMetGlnCysAsnTyrPheHisGlyValSerAsnCysThrP 866

    5048 AAGGGCGCATCAGTTTGACGTTCAGG 5073
           ||||||:::|||||||||||||||
     866 roGlyArgValSerLeuThrPheArg 874

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:R06531

seq_documentation_block:
ID   R06531 standard; protein; 992 AA.
XX
AC   R06531;
XX
DT   19-DEC-1990  (first entry)
XX
DE   Protein encoded by RNA 1 gene of CMV.
XX
KW   CMV; RNA 1.
XX
OS   Cucumber mosaic virus.
XX
PN   JP02167080-A.
XX
PD   27-JUN-1990.
XX
PF   19-DEC-1988;   88JP-0320015.
XX
PR   19-DEC-1988;   88JP-0320015.
XX
PA   (NOGY-) NOGYO SEIBUTSU IDEN.
XX
DR   WPI; 1990-241950/32.
DR   N-PSDB; Q05751.
XX
PT   Genome RNA 1 gene for prepn. of cucumber mosaic virus - codes
PT   protein contg. specified sequence of 994 amino acids used in prepn.
PT   of resistant plants.
XX
PS   Claim 1; Page 1; 10pp; Japanese.
XX
```

```
CC   The sequence was deduced from the DNA sequence which was prepd.
CC   from RNA isolated from CMV-O virus.  The gene can be used to
CC   prepare plants which are resistant to cucumber mosaic virus.
XX
SQ   Sequence   992 AA;


alignment_scores:
           Quality:  274.50            Length:    659
             Ratio:  0.850               Gaps:     25
 Percent Similarity:  49.014   Percent Identity: 22.155

alignment_block:
 US-09-301-906-3 x R06531   ..

 Align seg 1/1  to: R06531  from: 1  to: 992

   4696 TTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTC 4745
        :::::::::||||||:::   |||         ::::::|||   :::||
    434 TyrGluGlnLeuGlyLysMetTyrAspMetTrpAsnAlaSerSerIleSe 450

   4746 ATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACC 4795
        |:::   |||                    :::::::|||:::
    450 rLys...Trp.....................PheAlaAlaLeuThrArgA 459

   4796 CAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGC 4845
             ||||||            |||
    459 rgArgValPhePheSerSerAlaVal....................... 467

   4846 GTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTAT 4895
                |||||                 |||   ::::::
    468 .........HisAlaLeuProThrLeuArgProArgGluGluLysGl 481

   4896 CTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGG. 4944
         :::   :::::|||         ::::::|||:::   :::|||
    481 uPheLeuIleLysLeuSerThrPheValThrPheAsnGluGluCysSerP 498

   4945 ..AAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTT 4992
            ||||||:::         :::   ::::::|||:::   |||
    498 heAspGlyGlyGluGluTrpAspVal...IleSerSerAlaAlaTyrVal 513

   4993 CTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCAT 5042
                                              :::::::::::::
    514 ................................AlaThrGlnAlaVa 518

   5043 CGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTG 5092
        :   :::|||::::::      :::   :::|||   :::
    518 lThrAspGlyLysValLeuAlaAlaGlnLysAlaGluLysLeu....... 532

   5093 GTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142
          :::   ::::::   |||   :::      |||||||||      |||
    533 ..AlaGluLysLeuAlaGlnProValAspGluValSerAsp......Ser 546

   5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
        |||   ||||||::::::   |||:::
    547 ProGluValProSerSerThrProAsp...................... 555

   5193 CACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAG 5242

    555 .................................................. 555

   5243 TCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGG 5292
                  :::|||   ::::::      :::::::::   |||
    556 ...........AspThrAlaAspValCysGlyLysGluGlnGluValSer 568

   5293 AAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGA 5342
        :::   :::   |||   |||::::::
    569 GluLeuAspSerLeuSerAlaGlnThr...................... 577

   5343 CTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATG 5392
           :::|||:::|||   :::|||::::::   :::|||         |
    578 .ArgSerProIleThr...ArgValAlaGluArgAlaThr.........A 590

   5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC 5442
        ||::::::||||||      |||
    590 laMetLeuGluTyrAlaAlaTyr.......................... 597

   5443 AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGT 5492
                          :::::::::      ||||||:::::::::::
    598 ..............GluLysGlnLeuHisAspThrThrValSerAsnLe 609

   5493 TCTGAAAGTACCTAATCAAGTT..............TATGAATCACTAC 5527
        :   ::::::   |||::::::                  :::||||||
    609 uLysArgIleTrpAsnMetAlaGlyGlyAspAspLysArgAsnSerLeuG 626

   5528 CGGGC..............TTTCACGTTTACAAGTCGGGCACAGATCTC 5562
          |||               |||   :::|||   :::         :::
    626 luGlyAsnLeuLysPheValPheAspThrTyrPheThrValAspProMet 642

   5563 ATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGT 5612
        :::            :::   |||      :::|||   :::|||
    643 ValAsnIleHisPheSerThrGlyArgTrpMetArgProValProGluGl 659

   5613 ACTCATAGCTGAAAAAGGTATCTTTACCAAGGGC.........AAAGATG 5653
         ::::::   :::   |||         :::|||         |||:::
    659 yIleValTyrSerValGlyTyrAsnGluArgGlyLeuGlyProLysSerA 676

   5654 TCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATA 5703
              :::      ::::::::::::      :::|||::::::
    676 spGlyGluLeuPheIleValAsnSerGluCysValIleCysAsnSer... 691

   5704 CTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCG 5753
                 ::::::::::::   :::                |||
    692 ............GluSerLeuSerAlaValThrArgSerLeuGlnAlaPr 704

   5754 ATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATA 5803
        |||   :::::::::                              :
    704 oThrGlyThrIleSerGln.............................ValA 712

   5804 ATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTT 5853
        :::::      |||   ||||||||||||:::::   :::   |||
    712 spGlyValAlaGlyCysGlyLysThrThrAlaIleLysSerIlePhe... 727

   5854 AAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGA 5903
           :::|||:::::::          :::|||||||||      |||:::::
    728 ...GluProSerThrAspMetIleValThrAlaAsnLysLysSerAlaGl 743

   5904 GGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCA 5953
        :|||:::::::||||||:::   |||   :::   :::   |||:::
    743 nAspValArgMetAlaLeuPheLysSerSerAspSerLysGluAlaCysA 760

   5954 ACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGA 6003
           :::   |||::::::|||   ::::::|||
    760 laPheValArgThrAlaAspSerValLeuLeuAsnGluCys......Pro 774

   6004 ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTT 6053
           :::||||||||||||||||||||   ::::::|||   |||
    775 ThrValSerArgValLeuValAspGluValValLeuLeuHisPheGlyGl 791

   6054 ACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTG 6103
          |||         :::            |||   :::::::::   ||||
    791 nLeuCysAlaValMetSerLysLeuLysAlaValArgAlaIleCysPheG 808

   6104 GAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATG 6153
        |||||   :::||||||   |||   :::|||:::      |||   |||
    808 lyAspSerGluGlnIleAlaPheSerSerArgAspAlaSerPheAspMet 824

   6154 GATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAA 6203
           :::   ::::::|||   :::   |||   :::   |||
    825 ArgPheSerLysIleIleProAspGluThrSerAspAlaAspThr..... 839

   6204 ATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCG 6253
         ::::::|||   |||   ||||||      |||:::      |||::::
    840 .ThrPheArgSerProGlnAspValValProLeuValArgLeuMetAlaT 856

   6254 TA............AGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTT 6291
```

```
      ::           :::||||||::::::   |||   :::   |||
 856 hrLysAlaLeuProLysGlyThrHisSerLysTyrThrLysTrpValSer 872

6292 AGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGG 6341
      :::::|||         |||   ||||||:::::::::|||   |||::
 873 GlnSerLys.........ValLysArgSerValThrSerArgSerIleAl 886

6342 AACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGA 6391
      :::::::      |||         ::::::      |||:::   ||||
 886 aSerValThrLeuValAspLeuAspSerSerArgPheTyrIleThrMetT 903

6392 CCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAA 6441
      |||||   :::|||:::           ||||||   ::::::::|||
 903 hrGlnAlaAspLysAla...........SerLeuIleSerArgAlaLys 915

6442 GAA.......................ACACCAGTGATGACAGTGCA 6464
      |||                          :::   ||||||||
 916 GluMetAsnLeuProLysThrPheTrpAsnGluArgIleLysThrValHi 932

6465 TGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGA 6514
      ||||:::||||||   :::   :::   |||:::|||   |||   |||:
 932 sGluSerGlnGlyIleSerGluAspHisValThrLeuValArgLeuLysS 949

6515 AAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTG 6564
      :::::      :::||||||   :::      :::   ||||||:::|||
 949 erThrLysCysAspLeuPheLysGlnPheSerTyrCysLeuValAlaLeu 965

6565 TCGAGACACACACGCTCACTGGTTTAT 6591
      :::||||||      ::::::   |||
 966 ThrArgHisLysValThrPheArgTyr 974

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:R49657

seq_documentation_block:
ID   R49657 standard; Protein; 1704 AA.
XX
AC   R49657;
XX
DT   12-SEP-1994  (first entry)
XX
DE   Sequence of Heliothis armigera replicase encoded by RNA 1.
XX
KW   HaSV; RNA 1; small RNA virus; replicase.
XX
OS   Heliothis armigera stunt virus.
XX
PN   WO9404660-A.
XX
PD   03-MAR-1994.
XX
PF   13-AUG-1993;   93WO-AU00411.
XX
PR   14-AUG-1992;   92AU-0004081.
PR   08-JUL-1993;   93US-0089372.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (PACI-) PACIFIC SEEDS PTY LTD.
XX
PI   Christian PD,  Gordon KHJ,  Hanzlik TN;
XX
DR   WPI; 1994-083180/10.
DR   N-PSDB; Q58522.
XX
PT   Small RNA virus capable of infecting insect species, e.g.
PT   Heliothis - and transgenic plants contg. viral nucleic acid, for
PT   protection against insect pests
XX
PS   Disclosure; Figure 1; 183pp; English.
XX
CC   The inventors claim a virus comprising a genome hybridisable with
CC   the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC   those given in Figs 1 and 2 of the specification. Isolated proteins
CC   or polypeptide prepn. of the proteins or polypeptides derivable
CC   from the virus are also claimed.
CC   H. armigera larvae were raised and viral RNA was extracted. The virus
CC   RNAs were reverse transcribed into cDNA. Clone E3 represents 99.7%
CC   or RNA 1.(hr236 contains about 88% or RNA 2.)The full length clone
CC   of RNA 1 was completed using PCR. RNA 1 encodes a protein of mol.
CC   wt. 187,000 which is regarded as the replicase in view of its AA
CC   sequence similarity in certain limited regions to replicases of
CC   of other RNA viruses. The apparent mol. wt. of this protein upon
CC   in vitro translation of virus RNA and SDS-PAGE is 195,000. The
CC   sequence given in Figure 1 is in the same sense as the viral
CC   (positive-sense) RNA. There are other small ORFs at the 3' end,
CC   corresp. to the proteins P11a, P11b and P14.
XX
SQ   Sequence   1704 AA;


alignment_scores:
            Quality:  237.50              Length:    991
              Ratio:  0.571                Gaps:     50
 Percent Similarity:  41.978   Percent Identity: 21.191

alignment_block:
 US-09-301-906-3 x R49657   ..

 Align seg 1/1  to: R49657  from: 1  to: 1704

4273 GGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGA 4322
      |||         |||:::|||      ::::::::::   |||
   54 GlyThrLeuProProThrGlnHisProIleLeuAlaGlyHisGlnArgVa 70

4323 TATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAAC 4372
           :::|||   ::::::   ::::::::::::::   |||   :
   70 lAlaGluGluValLeuHisAsnPheAlaArgGlyArgSerThrValLeuG 87

4373 AAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAG 4422
      ::      |||     ::::::   :::|||      :::
   87 luIleGlyPro.....SerLeuHisSerAlaLeuLysLeuHisGlyAla 101

4423 AAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGG 4472
           ||||||::::::      :::|||
  102 ProAsnAlaProAlaAlaAspTyrHisGlyCysThrLysTyrGlyThrAr 118

4473 CAAGGCCGTCGCTCATGTGGCT......GAGAAAAAACAGGTACAAGTCG 4516
        :::      |||::::::      |||::::::::::|||   :::
  118 gAspGlySerArgHisIleThrAlaLeuGluSerArgSerValAlaThrG 135

4517 AGCAGCCCAAA.........CAGAGGAGTTTGACGATCAATGAAGGCAAG 4557
       :::|||:::          :::   ||||||      |||   |||
  135 lyArgProGluPheLysAlaAspAlaSerLeuLeuAlaAsn...GlyIle 150

4558 GCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGC........... 4596
      :::::::::   :::|||:::      :::   ||||||
  151 GlySerArgThrPheCysValAspGlyValGlySerCysAlaPheLysSe 167

4597 .......GGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCA 4639
             |||:::                  ::::::|||:::      :
  167 rArgValGlyIleAlaAsnHisSerLeuTyrAspValThrLeu...GluG 183

4640 GGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCG 4689
      ::::::::|||||||||
  183 luLeuAlaAsnAlaPheGlu............................. 189

4690 GTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCA 4739
                                                |||   |||   ||
  190 .................................AsnHisGlyLeuHi 194

4740 TGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTA.......... 4779
      |:::      ::::::   :::      |||::::::|||
  194 sMetValArgAlaPheMetHisMetProGluGluLeuLeuTyrMetAspA 211

4780 ..............................ACGGCAATTAAGTACCCA 4797
                                          :::|||:::   |||
  211 snValValAsnAlaGluLeuGlyTyrArgPheHisValIleGluGluPro 227
```


4798 AGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGT 4847
228 MetAlaValLysAspCysAlaPheGln.....GlyGlyAspLe 240
4848 ACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCT 4897
240 uArgLeuHisPheProGluLeuAspPheIleAsnGluSerGlnGluArgA 257
4898 TGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAG 4947
257 rgIleGluArgLeuAlaAlaArgGlySerTyrSerArgArgAla..... 271
4948 GGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGAC..... 4986
272 ValIlePheSerGlyAspAspAspTrpGl 281
4987 TATTTTCTTATG. 4998
281 yAspAlaTyrLeuHisAspPheHisThrTrpLeuProTyrLeuLeuValA 298
4999 CCTTGCGGTTTT.....CAAAGG 5016
298 rgAsnTyrProThrProPheGlyPheSerLeuHisIleGluValGlnArg 314
5017 ACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGAC 5066
315 ArgHisGlySerSerIleGlu..... 321
5067 GTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAG 5116
322 .LeuArgIleThrArgAlaProProGly...AspArgMetLeuAlaValV 337
5117 CC.....GGCGGCGTGTCGGATG..... 5134
337 alProArgThrSerGlnGly.LeuCysArgIleProAsnIlePheTyrTy 353
5135 AGAAGTCACCAGGTGTTCCAAACCAGCAACCACA 5168
353 rAlaAspAlaSerGlyThrGluHisLysThrIleLeuThrSerGlnHisL 370
5169 GAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTC 5218
370 ysValAsnMetLeu.LeuAsn..... 376
5219 TATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATA 5268
377 PheMetGlnThrArgProGluLysGluLeuValAspM 389
5269 TG.....GTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
389 etThrValLeuMetSerPheAlaArgAlaArgLeuArg...AlaIleVal 404
5301 GTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCC 5350
405 ValAla.....SerGlu.ValThrGluSerSerTrpAsnI 416
5351 CAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTG 5400
416 leSerProAlaAspLeuValArgThrValValSer..... 427
5401 GAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTT 5450
428 LeuTyrValLeuHisIleIleGluArgArgArgAlaAla.. 440
5451 GCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAG 5500
441 ValAlaValLysThrAlaLysAspAspValPheG 452
5501 TACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCG 5550
452 lyGluThrSerPheTrpGluSerLeuLys.....HisValPheGlySer 466
5551 GGCACAGATCTCATTTTT...CATTC 5573
467 CysCysGlyLeuArgAsnLeuLysGlyThrAspValValPheThrLysAr 483
5574 AACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGTACTC..... 5616
483 gValValAspLysTyrArgValHisSerLeuGlyAspIleIleCysAspV 500
5616 5616
500 alArgLeuSerProGluGlnValGlyPheLeuProSerArgValProPro 516
5617 ATAGCTGAAAAAGG 5630
517 AlaArgValPheHisAspArgGluGluLeuGluAlaLeuArgGluAlaGl 533
5631 TATCTTTACC.....A 5641
533 yCysTyrAsnGluArgProValProSerThrProProValGluGluProG 550
5642 AGGGCAAAGATGTCGAC..... 5658
550 lnGlyPheAspAlaAspLeuTrpHisAlaThrAlaAlaSerLeuProGlu 566
5659 GCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATAT 5702
567 TyrArgAlaThrLeuGlnAlaGlyLeuAsnThrAspValLysGlnLeuLy 583
5703 ACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTC 5752
583 sIleThrLeuGluAsnAlaLeuLysThrIleAspGlyLeuThrLeuSerP 600
5753 GATGC...GGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGC 5799
600 roValArgGlyLeu.....GluMet 606
5800 TATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTG..... 5841
607 TyrGluGlyProProGlySerGlyLysThrGlyThrLeuIleAlaAlaLe 623
5841 5841
623 uGluAlaAlaGlyGlyLysAlaLeuTyrValAlaProThrArgGluLeuA 640
5842 GACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATT 5880
640 rgGluAlaMetAspArgArgIleLysProProSerAlaSerAlaThr... 655
5881 ACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAG 5930
656 GlnHisValAlaLeuAlaIleLeuArgAr 665
5931 AGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGG 5980
665 g...AlaThrAlaGluGlyAlaProPheAlaThrValValIleAspGluC 681
5981 TGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAG 6030
681 ysPhe.....MetProLeuValTyrValAlaIle 691
6031 GTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGG 6080
692 ValHisAlaLeu.....SerPr 697
6081 CGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAA 6130
697 oSerSerArgIleValLeuValGlyAspValHisGlnIleGlyPheIleA 714
6131 ACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAG 6180
714 sp..... 714
6181 GAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTA.....GA 6224

```
                    :::   :::   |||   :::   ||||||      ||
     715 ...........PheGlnGlyThrSerAlaAsnMetProLeuValArgAs 727

    6225 TGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTT 6274
         ||||                    |||   :::   |||:::::|||
     727 pValValLysGlnCysArgArgArgThrPheAsnGlnThrLysArgCys. 743

    6275 ACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG 6324
           |||      ||||||:::                     :::||||||
     744 ..ProAlaAspValValAla.........ThrThrPhePheGlnSerLeu 756

    6325 TCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAAC..... 6369
                        |||:::      ||||||:::|||:::
     757 TyrProGlyCysThrThrThrSerGlyCysValAlaSerIleSerHisVa 773

    6370 .GCTGACGTGTAC.....................TTGTGCATGACCCAGT 6397
          |||      |||                     ||||||:::||||||
     773 lAlaProAspTyrArgAsnSerGlnAlaGlnThrLeuCysPheThrGlnG 790

    6398 TGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACA 6447
           |||||||||      :::::::::   :::|||
     790 luGluLysSer......ArgHisGlyAlaGluGly.............. 799

    6448 CCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGT 6497
            :::||||||||||||||||||||||||:::||||||::::::|||::
     800 ...AlaMetThrValHisGluAlaGlnGlyArgThrPheAlaSerVal1l 815

    6498 ATTG...TTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAAC 6544
         :|||   ::::::   :::   |||:::::::::|||:::::::::::::
     815 eLeuHisTyrAsnGlySerThrAlaGluGlnLysLeuLeuAlaGluLysS 832

    6545 CGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCC 6594
           |||:::||||||||:::::::||||||||:::   |||
     832 erHisLeuLeuValGlyIleThrArgHisThrAsnHisLeu......... 845

    6595 GCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGC 6644
                                              ||||||   |||
     846 .....................................TyrIleArgAsp.. 849

    6645 GTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGCTG 6694
            |||   :::      :::      |||:::      |||   |||
     850 ....ProThrGlyAspIleGluArgGlnLeuAsnHisSerAlaLysAlaG 865

    6695 GTTGCTTTCGAGGTATA 6711
              |||      |||
     865 luValPheThrAspIle 870

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y31381

seq_documentation_block:
ID   Y31381 standard; Protein; 1698 AA.
XX
AC   Y31381;
XX
DT   12-OCT-1999  (first entry)
XX
DE   HEV-US1 ORF1 protein.
XX
KW   Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW   vaccine; passive immunisation.
XX
OS   Hepatitis E virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 174
FT                   /label= unknown
FT                   /note= "encoded by ATR"
FT   Misc-difference 363
FT                   /label= unknown
FT                   /note= "encoded by RTG"
FT   Misc-difference 1088
FT                   /label= unknown
FT                   /note= "encoded by CRC"
FT   Misc-difference 1131
FT                   /label= unknown
FT                   /note= "encoded by YTC"
FT   Misc-difference 1217
FT                   /label= unknown
FT                   /note= "encoded by CRC"
FT   Misc-difference 1389
FT                   /label= unknown
FT                   /note= "encoded by GWC"
XX
PN   WO9919732-A1.
XX
PD   22-APR-1999.
XX
PF   15-OCT-1998;   98WO-US21941.
XX
PR   15-OCT-1997;   97US-0061199.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Dawson GJ,  Desai SM,  Erker JC,  Mushahwar IK,  Schlauder GG;
XX
DR   WPI; 1999-288017/24.
DR   N-PSDB; Z00195.
XX
PT   Detection of United States isolates of hepatitis E virus
XX
PS   Claim 6; Page 175-180; 260pp; English.
XX
CC   The invention provides a method for detecting a US (sub)type hepatitis E
CC   virus (US-HEV), or its naturally occurring variants in a sample by
CC   treatment with a binding partner specific for a marker of the virus, and
CC   then detecting any complex formed. The method is used to diagnose
CC   infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC   open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC   are useful in vaccines or for passive immunisation. The polypeptides are
CC   also used to raise specific antibodies (useful as immunoassay reagents).
CC   Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC   usual hybridisation and amplification assays for detecting infection. The
CC   present sequence represents a HEV-US1 ORF1 protein.
XX
SQ   Sequence   1698 AA;


alignment_scores:
            Quality:  188.50              Length:   1860
              Ratio:  0.274                 Gaps:     79
 Percent Similarity:  36.989   Percent Identity: 18.065

alignment_block:
 US-09-301-906-3 x Y31381   ..

 Align seg 1/1  to: Y31381  from: 1  to: 1698

    1375 TTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCT 1424
         |||||||:::      :::   |||   |||   :::|||:::
      28 PheLeuSerArgValGlnThrGluIleLeuIleAsnLeuMetGlnProAr 44

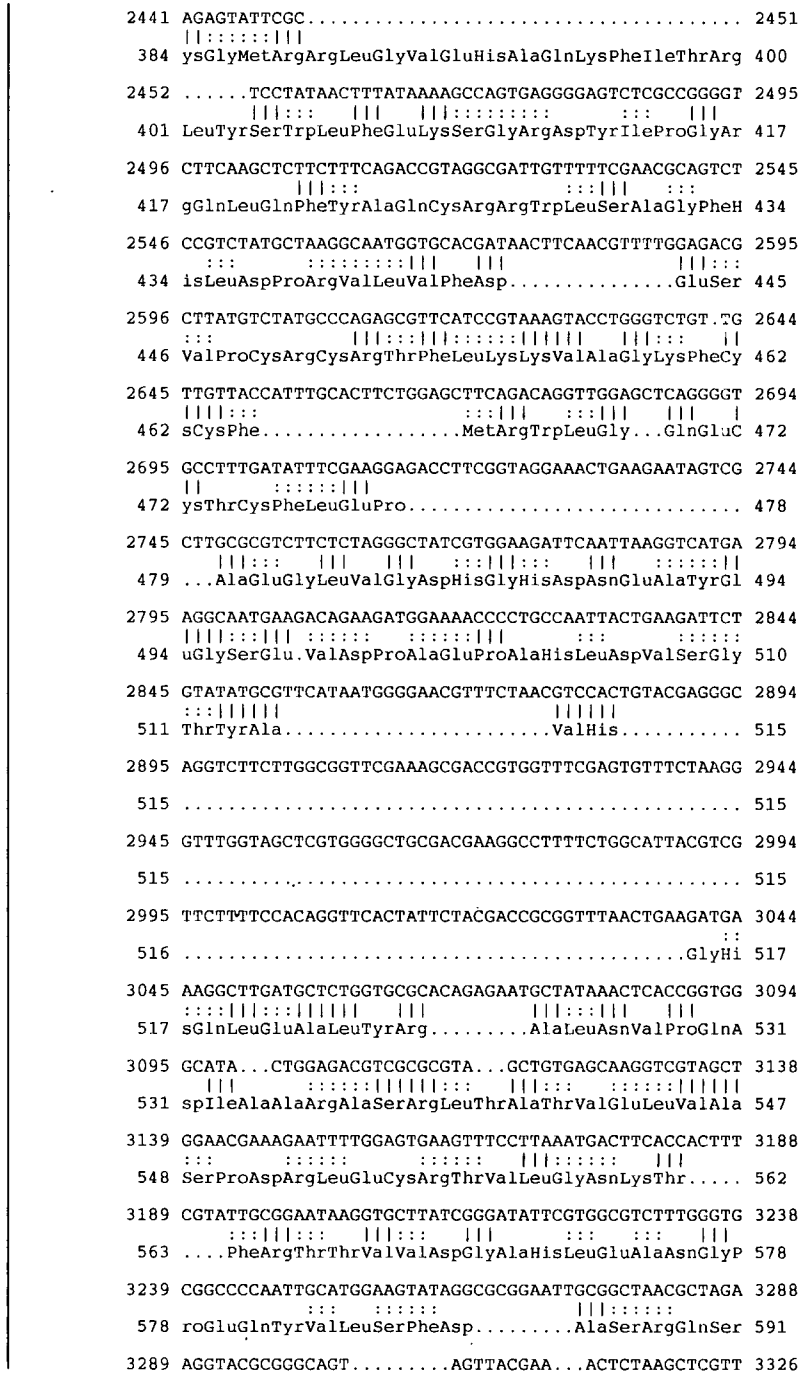
    1425 TTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATG 1474
          ::::::   |||      :::|||      :::||||||:::   ::::
      44 gGlnLeuValPheArgProGluValLeuTrpAsnHisProIleGlnArgV 61

    1475 CCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTG 1524
         ::::::::::   ::::::                |||:::|||:::
      61 alIleHisAsnGluLeuGluGlnTyrCysArgAlaArgAlaGlyArg... 76

    1525 TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCA 1574
         |||   :::::::::|||:::            |||   ::::::   ::
      77 Cys...LeuGluValGlyAla............HisProArgSerIleAs 88

    1575 TGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGC 1624
         :   |||   :::|||   :::      :::   ::::::   |||   |
      88 nAspAsnProAsnValLeuHisArgCysPhe...LeuArgProValGlyA 104
```





```
          |||||||||          :::||||||   |||   ::::::||
  592 MetGlyAlaGlySerHisSerLeuThrTyrGluLeuThrProAlaGlyLe 608

 3327 AAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTAT 3376
      |:::   :::   :::::::::            |||::::::
  608 uGlnValArgIleSerSerAsnGlyLeuAspCysThrAlaThrPheProP 625

 3377 CCGGTGGATCTTTAGTCGTGCGAAGAGGG.....TTTTCGTCG 3414
        ||||||:::      :::      |||          |||   |||
  625 roGlyGlyAlaProSerAlaAlaProGlyGluValAlaAlaPheCysSer 641

 3415 GCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGC 3464
      |||:::
  642 AlaLeu.....TyrAr 645

 3465 GTTGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAG 3514
               |||:::      :::::::::::::|||            |
  645 gTyrAsnArgPheThrGlnArgHisSerLeuThrGly.....G 658

 3515 GCATTTGGGCACATGCT.....CTT 3534
      ||:::|||   |||
  658 lyLeuTrpLeuHisProGluGlyLeuLeuGlyIleProProPheSer 674

 3535 CCACGGCACTTAATGTTT.....TTCTTTGGTTTAGGGACATT 3572
      |||   |||:::                  |||   |||   ||||||||
  675 ProGlyHisIleTrpGluSerAlaAsnProPheCysGlyGluGlyThrLe 691

 3573 GCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCA 3622
      |                  ::::::||||||:::   :::|||||||||:
  691 uTyr.....ThrArgThrTrpSerThrSerGlyPheSerS 703

 3623 ATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGC 3672
      ::
  703 er..... 703

 3673 TCATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTT 3722

  703 ..... 703

 3723 GCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGA 3772

  703 ..... 703

 3773 GGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGAC 3822
                                       ||||||:::
  704 .....AspPheSerProPro 708

 3823 GAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGA 3872
      |||:::|||:::|||:::              |||:::||||||
  709 GluAlaAlaAlaProAla.....MetAlaAlaThrProGlyLeuPr 722

 3873 TAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATG 3922
       :::      ||||||      ||||||:::      |||:::
  722 oHis......SerThrProProValSerAspIleTrpValLeu..... 734

 3923 TGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAG 3972

  734 ..... 734

 3973 ACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGC 4022
           |||   |||                    ||||||
  735 ......ProProProSer.....GluGluPheGl 742

 4023 CCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTA 4072
       :::||||||   :::          ||||||:::|||
  742 nValAspAlaAlaProValProProAlaProAspProAlaGly..... 756

 4073 CCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGAT 4122

  756 ..... 756

 4123 GTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGA 4172

  756 ..... 756

 4173 CGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAG 4222
                 ||||||      ||||||         |||
  757 .....LeuProGlyProValValLeuThrProPro..... 766

 4223 AAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAA 4272

  766 ..... 766

 4273 GGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGA 4322
                                          |||   :::
  767 .....ProProProProValHi 772

 4323 TATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAAC 4372
       |||::::::
  772 sLysProSerIle..... 776

 4373 AAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAG 4422

  776 ..... 776

 4423 AAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGG 4472
         |||   ||||||
  777 ...ProProProSer..... 780

 4473 CAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGC 4522

  780 ..... 780

 4523 CCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTT 4572
        ::::::|||   |||
  781 ..ArgAsnArgArgLeuLeu..... 786

 4573 TGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGA 4622

  786 ..... 786

 4623 AGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACT 4672

  786 ..... 786

 4673 TGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACC 4722
                                             ||||||
  787 .....TyrThr 788

 4723 TATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGA 4772
      |||       |||::   |||   :::|||
  789 TyrProAspGlyAlaLysValTyrAlaGly..... 798

 4773 TATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGAC...CACTGT.... 4815
                                  |||:::|||:::      |||
  799 .....SerLeuPheGluSerAspCysAspT 807

 4816 ..TTAGTGCAGAAG.....TACAAGATGGGTGGAGGCGTACCA 4851
        |||||||::                :::::   ||||||||||:::
  807 rpLeuValAsnAlaSerAsnProGlyHisArgProGlyGlyGlyLeuCys 823

 4852 .....TTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCC 4892
            :::::          |||      ||||||::::::
  824 HisAlaPheTyrGlnArgPheProGluAlaPheTyrProThrGlu..... 838

 4893 TATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCA 4942

  838 ..... 838

 4943 GGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTT 4992
                                                    |||
  839 .....Phe 839

 4993 CTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCAT 5042
      :::|||      |||:::
  840 IleMetArgGluGlyLeu..... 845
```

```
5043 CGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTG 5092

 845 ..................................................  845

5093 GTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142

 845 ..................................................  845

5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
                                          |||:::   |||::
 846 ..................................AlaAlaTyrThrLe 850

5193 CACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAG 5242
     :||||||:::
 850 uThrProArg..................................... 853

5243 TCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGG 5292

 853 ..................................................  853

5293 AAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATG...GCTCTTGAACC 5339
                                  |||::::::   |||:::   ||
 854 ........................ProIleIleHisAlaValAlaPr 861

5340 TGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAG 5389
     |||||||
 861 oAspTyr..................................... 863

5390 ATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACA 5439
         |||   :::     |||   |||           |||   :::
 864 ..ArgValGluGlnAsnProLysArgLeu.....GluAlaAla 875

5440 TACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGG 5489
     ||||||           |||:::   |||:::
 876 TyrArgGluThrCysSerArgArgGlyThrAlaAla..... 887

5490 TGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACG 5539
                                      |||   |||     :
 888 ........................TyrProLeuLeuGlySerGlyI 895

5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTG 5589
     ::|||:::     ::::::|||   |||   :::   :::   :::
 895 leTyrGlnValProValSerLeuSerPheAspAlaTrpGluArgAsnHis 911

5590 CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTAC 5639
     |||     |||     |||:::                 :::   |||
 912 ArgProGlyAspGluLeuTyrLeuThrGluProAlaAlaAsnTrpPheGl 928

5640 CAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCG 5689
         :::|||   ::::::   |||:::::::::   :::
 928 uAlaAsnLysProAlaGlnProValLeuThrIleThrGlu..... 941

5690 TATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCA 5739
         |||     :::   :::   |||:::::::::   :::|||
 942 .....AspThrAlaArgThrAlaAsnLeuAlaLeuGluIleAspAlaAla 956

5740 CTGAAAGTCGCTCGA...TGCGGCATGGTGGGCGAATCGTTTAAGTCCTT 5786
        :::|||:::|||   |||:::           |||
 957 ThrGluValGlyArgAlaCysAlaGlyCysThrIleSerProGlyIleVa 973

5787 CGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGT 5836
      :::|||:::         :::   ||||||:::||||||:::   ::::
 973 lHisTyrGlnPheThrAlaGlyValProGlySerGlyLysSerArgSerI 990

5837 TAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCT 5886
     ::                     :::   :::   :::   :::
 990 le.......................GlnGlnGlyAspValAspVal 997

5887 AATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCC 5936
        |||      :::::::::::::::   :::   ::::::|||
 998 ValValValProThrArgGluLeuArgAsnSerTrpArgArgArg..... 1012

5937 GAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAA 5986
               |||:::   :::   |||      ::::::      |||:
1013 .........GlyPheAlaAlaPheThrProHisThrAlaAlaArgValT 1026

5987 ACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCAC 6036
     ::                :::   :::|||||:::::|||||:::
1026 hr..............IleGlyArgArgValValIleAspGluAlaPro 1037

6037 ATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTC 6086
        :::         ||||||   |||   :::             |||||
1038 SerLeuProProHisLeuLeuLeuLeuHisMet......GlnArgAlaSe 1052

6087 GGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGG 6136
     |:::      ::::::||||||   ||||||||||||   |||:::   |
1052 rSerValHisLeuLeuGlyAspProAsnGlnIleProAlaIleAspPheG 1069

6137 AGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAG...AAGGAA 6183
     ||                      |||   :::||||||      :::
1069 luHis..................AlaGlyLeuValProAlaIleArgPro 1079

6184 AGCGTTGTATACACTTCT.........AAATCGTACAGGTGTCCGTTAGA 6224
     :::::::::   ||||||          ::::::||||||||||   ||
1080 GluLeuAlaProThrSerTrpTrp***ValThrHisArgCysProAlaAs 1096

6225 TGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTT 6274
     |||||||   |||               :::||||||         :::|
1096 pValCysGluLeu..............IleArgGly.........AlaT 1105

6275 ACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG 6324
     |||||:::      ::::::::::::          |||:::|||||||||
1105 yrProLysIleGlnThrThrSerArg.........ValLeuArgSerLeu 1118

6325 ......TCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAA 6368
           ::::::      ||||||
1119 PheTrpAsnGluProAlaIleGlyGlnLys..................... 1128

6369 CGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGA 6418
                  |||      ||||||
1129 .............LeuVal***ThrGln..................... 1133

6419 GGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAA 6468
                  :::|||:::   :::      ::::::||||||||||||
1134 ........AlaAlaLysAlaAlaAsnProGlyAlaIleThrValHisGlu 1147

6469 GCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGC 6518
     |||||||||   ||||||:::::::::::::::::::   |||      ||
1148 AlaGlnGlyAlaThrPheThrGluThrThrIleIleAlaThrAlaAspAl 1164

6519 CGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGA 6568
     |         :::   :::::::::   |||   :::|||:::|||:::|
1164 aArgGlyLeuIleGlnSerSerArgAlaHisAlaIleValAlaLeuThrA 1181

6569 GACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGAT 6618
     ||||||||:::::::   |||                :::   |||   :::
1181 rgHisThrGluLysCysValIleLeuAspAlaProGlyLeuLeuArgGlu 1197

6619 AAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCAGTATCCGACGC 6668
        ||||||                                :::||||||::
1198 ...ValGly..............................IleSerAspVa 1203

6669 TTTGCTTCACACGTTCGCCCCGGCTGGT 6696
     :::::::::::::|||      ||||||
1203 lIleValAsnAsnPhePheLeuAlaGly 1212
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:Y31384

seq_documentation_block:
ID Y31384 standard; Protein; 1708 AA.
XX
AC Y31384;
XX
DT 12-OCT-1999 (first entry)
XX

```
DE   HEV-US2 ORF1 protein.
XX
KW   Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW   vaccine; passive immunisation.
XX
OS   Hepatitis E virus.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 322
FT                   /label= unknown
FT                   /note= "encoded by CRG"
FT   Misc-difference 331
FT                   /label= unknown
FT                   /note= "encoded by RAC"
FT   Misc-difference 445
FT                   /label= unknown
FT                   /note= "encoded by RAC"
FT   Misc-difference 448
FT                   /label= unknown
FT                   /note= "encoded by RTG"
FT   Misc-difference 634
FT                   /label= unknown
FT                   /note= "encoded by CCY"
FT   Misc-difference 646
FT                   /label= unknown
FT                   /note= "encoded by GCS"
FT   Misc-difference 811
FT                   /label= unknown
FT                   /note= "encoded by TKT"
FT   Misc-difference 1533
FT                   /label= unknown
FT                   /note= "encoded by TAY"
FT   Misc-difference 1578
FT                   /label= unknown
FT                   /note= "encoded by RGC"
FT   Misc-difference 1691
FT                   /label= unknown
FT                   /note= "encoded by GAR"
XX
PN   WO9919732-A1.
XX
PD   22-APR-1999.
XX
PF   15-OCT-1998;   98WO-US21941.
XX
PR   15-OCT-1997;   97US-0061199.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Dawson GJ,  Desai SM,  Erker JC,  Mushahwar IK,  Schlauder GG;
XX
DR   WPI; 1999-288017/24.
DR   N-PSDB; Z00267.
XX
PT   Detection of United States isolates of hepatitis E virus
XX
PS   Claim 6; Page 216-221; 260pp; English.
XX
CC   The invention provides a method for detecting a US (sub)type hepatitis E
CC   virus (US-HEV), or its naturally occurring variants in a sample by
CC   treatment with a binding partner specific for a marker of the virus, and
CC   then detecting any complex formed. The method is used to diagnose
CC   infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC   open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC   are useful in vaccines or for passive immunisation. The polypeptides are
CC   also used to raise specific antibodies (useful as immunoassay reagents).
CC   Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC   usual hybridisation and amplification assays for detecting infection. The
CC   present sequence represents a HEV-US2 ORF1 protein.
XX
SQ   Sequence   1708 AA;
```

```
alignment_scores:
          Quality:  185.00              Length:   1864
            Ratio:  0.272                Gaps:     77
 Percent Similarity:  36.534   Percent Identity: 17.918

alignment_block:
 US-09-301-906-3 x Y31384    ..

 Align seg 1/1  to: Y31384  from: 1  to: 1708

    1375 TTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCT 1424
         |||||||:::      :::   |||   |||   :::|||:::
      37 PheLeuSerArgValGlnThrGluIleLeuIleAsnLeuMetGlnProAr 53

    1425 TTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATG 1474
          ::::::   |||      :::|||      :::||||||:::   ::::
      53 gGlnLeuValPheArgProGluValLeuTrpAsnHisProIleGlnArgV 70

    1475 CCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTG 1524
         ::::::::::   ::::::                  |||:::|||:::
      70 alIleHisAsnGluLeuGluGlnTyrCysArgAlaArgAlaGlyArg... 85

    1525 TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCA 1574
         |||   :::::::::|||:::            |||   ::::::   ::
      86 Cys...LeuGluValGlyAla...........HisProArgSerIleAs 97

    1575 TGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGC 1624
         :   |||   :::|||  :::      :::   ::::::   |||   |
      97 nAspAsnProAsnValLeuHisArgCysPhe...LeuArgProValGlyA 113

    1625 GGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTAC 1674
         ||   :::::::::            |||   |||      :::
     113 rgAspValGlnArgTrpTyrSerAlaProThrArgGlyProAlaAlaAsn 129

    1675 GTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTG 1724
               :::   |||                  :::   :::|||||
     130 CysArgArgSerAlaLeuArgGlyLeuProProValAspArgThrTyrCy 146

    1725 TAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTT...A 1771
         |           ::::::|||         |||:::::|||      :
     146 sPheAspGlyPheSerArgCysAlaPheAlaAlaGluThrGlyValAlaL 163

    1772 TGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGAT 1821
         ::   :::::::::|||:::   |||   :::|||||||||||||||
     163 euTyrSerLeuHisAspLeuTrpProAlaAspValAlaGluAlaMetAla 179

    1822 AAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTT 1871
         :::   |||         :::      ||||||   :::|||   |||::
     180 ArgHisGlyMetThrArgLeuTyrAlaAlaLeuHisLeuProProGluVa 196

    1872 GTTGTACGGTAACGGTGAAGTT...........TACTTGGAAGAACTCG 1909
         :|||         |||                 ||||||        |
     196 lLeuLeuProProGlyThrTyrHisThrThrSerTyrLeuLeuIleHisA 213

    1910 AT.........ACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAAT 1950
         ||         :::::::|||      |||||||||      |||
     213 spGlyAsnArgAlaValValThrTyrGluGlyAspThrSerAla...... 227

    1951 GTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGG 2000
                           |||:::|||:::   |||   :::
     228 ..................GlyTyrAsnHisAspValSerIleLeuArg.. 237

    2001 GTTTTTCACCTTTTCTTATGTACGCACTTCGTCC......GGGAACGTGT 2044
                        :::::::::||||||::::::      |||:::
     238 .............AlaTrpIleArgThrThrLysIleValGlyAsp...H 249

    2045 TTAAGCTAGAGTATGAGGGATACCGTTGT...GGTTACCATCATCTCACT 2091
              |||      |||      |||:::   |||   |||   :::
     249 isProLeuValIleGluArgValArgAlaIleGlyCysHisPheValLeu 265

    2092 ATGTGTAGGGCTCAGAAGTCACCT........................... 2115
         :::      |||      :::|||
     266 LeuLeuThrAlaAlaProGluProSerProMetProTyrValProTyrPr 282
```

```
2116 ....GGAACTGAGGTTACGTATAGG..... 2136
          :::|||||||||      |||
 282 oArgSerThrGluValTyrValArgSerIlePheGlyProGlyGlySerP 299

2137 ..TCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTC......ATA 2178
       ||||||   ||||||      :::   :::   :::|||      :::
 299 roSerLeuPheProSerAlaCysSerThrLysSerThrPheHisAlaVal 315

2179 CCTGTT...GTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCGA 2225
     ||||||   :::      :::   :::   |||   :::   :::   ||
 316 ProValHisIleTrpAsp***LeuMetLeuPheGlyAlaThrLeu***As 332

2226 TTCGGACTTT...GTCGACAGGATCTATTCCTACGCGCTCAACACTATAG 2272
     |:::   |||      :::|||:::   :::|||
 332 pGlnAlaPheCysCysSerArgLeuMetThrTyr..... 343

2273 GGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCG 2322
                                        :::|||:::
 344 .....LeuArgGly 346

2323 CAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGA 2372
        :::      |||   :::|||:::   |||:::::::::      ::
 347 IleSerTyrLysValThrValGlyAlaLeuValAlaAsnGluGlyTrpAs 363

2373 TATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCT..... 2409
     :   |||   |||   :::   :::::::::::|||
 363 nAlaSerGluAspAlaLeuThrAlaValIleThrAlaAlaTyrLeuThrI 380

2410 .....GTTATGGCTCAGGCGATTAAGGATAGGGCGAAG 2442
                  :::   :::|||||||||:::      :::
 380 leCysHisGlnArgTyrLeuArgThrGlnAlaIleSerLysGlyMetArg 396

2443 AGTATTCGC.....TCCTA 2456
        ::::::      |||::
 397 ArgLeuGluValGluHisAlaGlnLysPheIleThrArgLeuTyrSerTr 413

2457 TAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCT 2506
     :   |||   |||:::::::::      :::   |||           |
 413 pLeuPheGluLysSerGlyArgAspTyrIleProGlyArgGlnLeuGlnP 430

2507 TCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCT 2556
     ||:::                  :::|||   :::      :::
 430 heTyrAlaGlnCysArgArgTrpLeuSerAlaGlyPheHisLeu***Pro 446

2557 AAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGCTTATGTCTAT 2606
     ::::::::::|||   |||              |||:::::::
 447 Arg***LeuValPheAsp.....GluSerValProCysAr 458

2607 GCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGT.TGTTGTTACCATT 2655
        |||:::|||:::::::||||||   |||:::   ||||||:::
 458 gCysArgThrPheLeuLysLysValAlaGlyLysPheCysCysPhe.... 473

2656 TGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATAT 2705
                  :::|||   :::|||   |||   |||     :::
 474 .....MetArgTrpLeuGly...GlnGluCysThrCysPhe 484

2706 TTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCT 2755
     :::|||                                  |||:::
 485 LeuGluPro.....AlaGluGl 490

2756 TCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAG 2805
      |||   |||   :::|||:::   |||   ::::::||||||:::|||
 490 yLeuValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerGlu. 506

2806 ACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTT 2855
     ::::::   ::::::|||     :::      :::::::::||||||
 507 ValAspProAlaGluProAlaHisLeuAspValSerGlyThrTyrAla.. 522

2856 CATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTG 2905
                          ||||||
 523 .....ValHis..... 524
```

```
2906 GCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCT 2955

 524 ..... 524

2956 CGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCAC 3005

 524 ..... 524

3006 AGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATG 3055
                                        ::::::|||:::|
 525 .....GlyHisGlnLeuGluA 530

3056 CTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATA...CTG 3102
     |||||   |||         |||:::|||   |||      |||
 530 laLeuTyrArg.....AlaLeuAsnValProHisAspIleAlaAla 543

3103 GAGACGTCGCGCGTA...GCTGTGAGCAAGGTCGTAGCTGGAACGAAAGA 3149
     ::::::||||||:::   |||:::   ::::::||||||:::      ::
 544 ArgAlaSerArgLeuThrAlaThrValGluLeuValAlaSerProAspAr 560

3150 ATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGA 3199
     ::::      ::::::   |||:::::::   |||         :::||.:
 560 gLeuGluCysArgThrValLeuGlyAsnLysThr.....PheArgT 574

3200 ATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATT 3249
     ::   |||:::   |||       :::   :::   |||
 574 hrThrValValAspGlyAlaHisLeuGluAlaAsnGlyProGluGluTyr 590

3250 GCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGG 3299
     :::   ::::::            |||::::::            |||||
 591 ValLeuSerPheAsp.....AlaSerArgGlnSerMetGlyAlaGl 604

3300 CAGT.....AGTTACGAA...ACTCTAAGCTCGTTAAGTTCACAAG 3337
     ||||         :::|||||||   |||   ::::::|||:::   :::
 604 ySerHisSerLeuThrTyrGluLeuThrProAlaGlyLeuGlnValLysI 621

3338 CCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCT 3387
       :::::::::            |||::::::      :::|||||||:::
 621 leSerSerAsnGlyLeuAspCysThrAlaThrPhePro***GlyGlyAla 637

3388 TTAGTCGTGCGAAGAGGG.....TTTTCGTCGGCGGTGACCGT 3425
          :::      |||            |||   ||||||:::
 638 ProSerAlaAlaProGlyGluVal***AlaPheCysSerAlaLeu..... 652

3426 CACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGT 3475
                                                      |
 653 .....TyrArgTyrAsnArgP 658

3476 TTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCA 3525
     ||:::      :::::::::::|||            |||:::|||
 658 heThrGlnArgHisSerLeuThrGly.....GlyLeuTrpLeu 670

3526 CATGCT.....CTTCCACGGCACTT 3545
     |||                  |||   |||::
 671 HisProGluGlyLeuLeuGlyIleProProPheSerProGlyHisIl 687

3546 AATGTTT.....TTCTTTGGTTTAGGGACATTGCTTGGGGCGA 3583
     :              |||   |||   |||||||||
 687 eTrpGluSerAlaAsnProPheCysGlyGluGlyThrLeuTyr..... 701

3584 GGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGC 3633
            ::::::||||||:::   :::|||||||||:::
 702 .....ThrArgThrTrpSerThrSerGlyPheSerSer..... 712

3634 GCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACT 3683

 712 ..... 712

3684 GCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACA 3733

 712 ..... 712

3734 CCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTAC 3783
```

```
 712 ..................................................  712

3784 GATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGG 3833
                           |||||:::     |||:::|||::
 713 ......................AspPheSerProProGluAlaAlaAl 721

3834 TCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGAT 3883
     :|||:::::::::         ::::::|||
 721 aProAlaSerAlaAlaAlaProGlyLeuPro.............TyrP 733

3884 CCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTG 3933
       |||     ||||||:::     |||:::
 733 roThrProProValSerAspIleTrpValLeu................. 743

3934 GGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACC 3983
                                                     ||
 744 ..............................................Pr 744

3984 TCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGCAG 4033
     |     :::   ::::::|||:::       :::|||||||:::
 744 oProProSerGluGluSerHisValAspAlaAlaSerValProSerValP 761

4034 CTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACGCT 4083
                        :::|||:::|||
 761 ro.................GluProAlaGly..................... 765

4084 ATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCA 4133

 765 ..................................................  765

4134 ACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCAAAGGGT 4183

 765 ..................................................  765

4184 TGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGGGGAAGA 4233

 765 ..................................................  765

4234 GCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGACAGGGA 4283

 765 ..................................................  765

4284 GCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGG 4333
       ::   |||   :::|||   :::|||          |||   |
 766 ....LeuThrSerProIleValLeuThrProProProProProProProV 781

4334 TCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACCA 4383
     ||
 781 al................................................. 781

4384 TTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCCCC 4433
                                        :::||||||||||
 782 ..................................ArgLysProAlaTh 786

4434 AAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCG 4483
      |||
 786 rSer.............................................. 787

4484 CTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGG 4533
                                           |||:::   |||
 788 ..............................ProProProArgThrArg 793

4534 AGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAG 4583
       |||
 794 ArgLeuLeu.......................................... 796

4584 AACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCG 4633

 796 ..................................................  796

4634 CCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGG 4683
```

```
 796 ..................................................  796

4684 AGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGG 4733
                                        |||||||||      ||
 797 ..............................TyrThrTyrProAspGl 802

4734 TAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAG...GATATCTTAA 4780
     |:::   |||   :::|||                :::   |||   |||:
 802 yAlaLysValTyrAlaGlySerLeu***GluSerAspCysAspTrpLeuV 819

4781 CGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTAC 4830
     ::       :::   |||:::                            :::
 819 alAsnAlaSerAsnProGly.........................His 826

4831 AAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCC 4880
     :::   ||||||||||:::       ||||||      ::::::   :::||
 827 ArgProGlyGlyGlyLeu...CysHisAlaPheTyrGlnArg...PhePr 841

4881 ATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCT 4930
     |                                      :::|||   :::|
 841 o................................GluAlaPheTyrS 846

4931 CGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCG 4980
     |||||:::
 846 erThrGlu..................................... 848

4981 GGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTC 5030
               |||:::|||      |||:::
 849 .........PheIleMetArgGluGlyLeu..................... 855

5031 AGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTC 5080

 855 ..................................................  855

5081 GGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCG 5130

 855 ..................................................  855

5131 GATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGC 5180
                                                     ||
 856 ..............................................Al 856

5181 TACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAA 5230
     |:::   |||:::||||||:::
 856 aAlaTyrThrLeuThrProArg.......................... 863

5231 GTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAA 5280

 863 ..................................................  863

5281 GATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATG.. 5328
                                           |||::::::
 864 ..................................ProIleIleHi 867

5329 .GCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCG 5377
      |||:::   |||||||||
 867 sAlaValAlaProAspTyr.............................. 873

5378 GGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGC 5427
                    |||   :::      |||   |||
 874 ..............ArgValGluGlnAsnProLysArgLeu..... 882

5428 ATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCAC 5477
        |||   :::||||||               |||:::   |||:::
 883 ...GluAlaAlaTyrArgGluThrCysSerArgArgGlyThrAlaAla.. 897

5478 TACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTAC 5527
                                                    |||
 898 ..................................TyrProLeuL 901

5528 CGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACA 5577
       |||      :::|||:::      ::::::|||   |||   :::
 901 euGlySerGlyIleTyrGlnValProValSerLeuSerPheAspAlaTrp 917
```

```
5578 CAAGACGGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAA 5627
     :::   :::   |||      |||      |||:::
 918 GluArgAsnHisArgProGlyAspGluLeuTyrLeuThrGluProAlaAl 934

5628 AGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCG 5677
      :::   |||      :::|||   ::::::   ::::::::::::   :
 934 aAlaTrpPheGluAlaAsnLysProAlaGlnProAlaLeuThrIleThrG 951

5678 ACAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAAT 5727
     ::               |||         :::   :::   |||::::::
 951 lu..............AspThrAlaArgThrAlaAsnLeuAlaLeuGlu 962

5728 TTGATAGGTGCACTGAAAGTCGCTCGA...TGCGGCATGGTGGGCGAATC 5774
     :::   :::|||   :::|||:::|||   |||:::            ||
 963 IleAspAlaAlaThrGluValGlyArgAlaCysAlaGlyCysThrIleSe 979

5775 GTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTA 5824
     |           :::|||:::        :::   ||||||:::||||
 979 rProGlyIleValHisTyrGlnPheThrAlaGlyValProGlySerGlyL 996

5825 AGACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCC 5874
     ||:::   ::::::                              :::   :::
 996 ysSerArgSerIle.........................GlnGlnGly 1003

5875 ACCATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAA 5924
        :::   :::   |||      ::::::::::::::   :::   ::
1004 AspValAspValValValValProThrArgGluLeuArgAsnSerTrpAr 1020

5925 GAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACT 5974
     ::::|||              |||:::   :::   |||      ::::
1020 gArgArg..............GlyPheAlaAlaPheThrProHisThrA 1032

5975 CCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTG 6024
     ::      |||:::               :::   :::||||||:::::
1032 laAlaArgValThr..............IleGlyArgArgValValIle 1043

6025 GATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGC 6074
     ||||||:::      :::         ||||||   |||   :::
1044 AspGluAlaProSerLeuProProHisLeuLeuLeuLeuHisMet..... 1058

6075 AACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCAT 6124
          ||||||:::      ::::::||||||   ||||||||||||
1059 .GlnArgAlaSerSerValHisLeuLeuGlyAspProAsnGlnIleProA 1075

6125 TCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCA 6174
       |||:::   |||                    |||   :::||||||
1075 laIleAspPheGluHis.................AlaGlyLeuValPro 1085

6175 AAG...AAGGAAAGCGTTGTATACACTTCT.........AAATCGTACAG 6212
           :::   :::::::::   ||||||            ::::::||
1086 AlaIleArgProGluLeuAlaProThrSerTrpTrpHisValThrHisAr 1102

6213 GTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAA 6262
     |||||||   |||||||||   |||              :::||||||
1102 gCysProAlaAspValCysGluLeu..............IleArgGly. 1113

6263 CGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTA 6312
             :::|||||||:::      ::::::::::::         |||
1114 ........AlaTyrProLysIleGlnThrThrSerArg.........Val 1124

6313 GTAAGATCGCTG......TCCAAAAGGCCAATTGGAACCACTGATGACGT 6356
     :::|||||||||      ::::::      ||||||
1125 LeuArgSerLeuPheTrpAsnGluProAlaIleGlyGlnLys........ 1138

6357 AGCTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGT 6406
                           |||   :::||||||
1139 .......................LeuValPheThrGln.......... 1143

6407 CGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATG 6456
                          :::|||:::   :::      ::::::
1144 ..................AlaAlaLysAlaAlaAsnProGlyAlaIle 1153

6457 ACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAG 6506
     ||||||||||||||||||||   ||||||:::::::::::::::::
1154 ThrValHisGluAlaGlnGlyAlaThrPheThrGluThrThrIleIleAl 1170

6507 GACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTG 6556
      |||      |||         :::   :::::::::   |||   :::|
1170 aThrAlaAspAlaArgGlyLeuIleGlnSerSerArgAlaHisAlaIleV 1187

6557 TTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCA 6606
     ||:::|||:::||||||||||::::::   |||              :::
1187 alAlaLeuThrArgHisThrGluLysCysValIleLeuAspAlaProGly 1203

6607 AAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATC 6656
        |||   :::   ||||||
1204 LeuLeuArgGlu...ValGly............................ 1209

6657 AGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGCTGGT 6696
      :::||||||:::::::::::::::|||      ||||||
1210 .IleSerAspValIleValAsnAsnPhePheLeuAlaGly 1222
```

```
OM of: US-09-301-906-3 to: Issued_Patents_AA:*   out_format : pfs

Date: Jan 31, 2001  3:35 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09301906/runat_31012001_153034_9592/app_query.fasta_1.6881
 -DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09301906_@CGN1_1_208 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-3
Query length: 6714
Database: Issued_Patents_AA:*
Database sequences: 174772
Database length: 17957048
Search time (sec): 148.720000



score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-770-544-2 +  5247.00 8832.93   0.0       1390  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-09-081-320-3 +   290.00  428.86  7.1e-18    2161  !  S
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-840-316-1 +   179.50  245.18  1.5e-07    1693  !  S
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-08849A-1 +   179.50  245.18  1.5e-07    1693
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-08849-1 +   179.50  245.18  1.5e-07    1693
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-325-267A-2 -   178.00  244.04  2.0e-07    1537  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-478-507-7 +   169.50  228.24  1.4e-06    1693  !  S
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-616-844-40 -   160.50  214.92  8.5e-06    1481  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-599-654-40 -   160.50  214.92  8.5e-06    1481  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-868A-40 -   160.50  214.92  8.5e-06    1481  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-423A-40 -   160.50  214.92  8.5e-06    1481  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-496-40 -   160.50  214.92  8.5e-06    1481  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-682-517-15 +   150.50  200.76  6.4e-05    1222  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-682-517-9 +   150.50  200.41  6.5e-05    1252  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-362-525-22 -   149.50  203.57  6.1e-05     894  !  S
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-971-692-15 -   149.50  203.57  6.1e-05     894  !  S
/cgn2_6/ptodata/2/iaa/backfiles1.pep:5183745-6 +   149.50  193.52  0.0001     1794  ! Pate
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-551-356-2 +   145.00  181.42  0.0004     2446  !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-12687-2 +   145.00  181.42  0.0004     2446
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6 +   143.50  182.93  0.0004     1848
/cgn2_6/ptodata/2/iaa/backfiles1.pep:5183745-2 +   142.50  184.35  0.0004     1489  ! Pate
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-283-857-1 +   140.00  173.69  0.0011     2324  !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-09819-1 +   140.00  173.69  0.0011     2324
/cgn2_6/ptodata/2/iaa/backfiles1.pep:5455158-1 +   140.00  173.67  0.0011     2327  ! Pate
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-409-995-4 +   138.50  173.97  0.0013     1912  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-685-467-4 +   138.50  173.97  0.0013     1912  !  S
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-375-709-11 +   136.50  165.30  0.0027     2756  !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-752-929-11 +   136.50  165.30  0.0027     2756  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-09-090-793-7 +   136.50  165.30  0.0027     2756  !  S
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-232-537-2 +   135.50  174.43  0.0017     1302  !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-10661A-5 +   135.00  169.72  0.0024     1702
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-325-267A-4 -   134.50  178.69  0.0015      862  !
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-10661A-3 +   134.50  170.30  0.0025     1541
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-446-137B-2 +   134.00  175.31  0.0020     1027  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-821-355A-7 -   133.50  159.13  0.0054     2973  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-003-687A-7 -   133.50  159.13  0.0054     2973  !
/cgn2_6/ptodata/2/iaa/6_COMB.pep:US-09-136-605-7 -   133.50  159.13  0.0054     2973  !  S
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-614-377A-7 +   132.50  172.79  0.0027     1026  !
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-016-366A-12 +   132.50  160.61  0.0056     2386  !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-741-940-2 -   132.00  157.23  0.0073     2843  !
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-289-548A-2 -   132.00  157.23  0.0073     2843
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-452-654-2 -   132.00  157.23  0.0073     2843
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-370-235A-2 -   132.00  157.23  0.0073     2843
/cgn2_6/ptodata/2/iaa/backfiles1.pep:5180808-2 -   131.50  158.77  0.0070     2409  ! P
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-804-227C-5 +   131.00  163.73  0.0056     1611

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-770-544-2

seq_documentation_block:
; Sequence 2, Application US/08770544
; Patent No. 5907085
;  GENERAL INFORMATION:
;    APPLICANT:  Gonsalves, Dennis .
;    APPLICANT:  Ling, Kai-Shu
;    TITLE OF INVENTION:  GRAPEVINE LEAFROLL VIRUS PROTEINS AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/770,544
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60009008
;      FILING DATE:  21-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/621
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1390 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-770-544-2


alignment_scores:
            Quality: 5247.00              Length:   1423
              Ratio:  4.214                Gaps:     17
 Percent Similarity:  87.491   Percent Identity: 74.139 .

alignment_block:
 US-09-301-906-3 x US-08-770-544-2   ..

 Align seg 1/1  to: US-08-770-544-2  from: 1  to: 1390

    2542 GTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGA 2591
         |||||||:::|||||||||:::::::::::::|||||||||||||:::|||||
       1 ValSerThrTyrAlaLysSerValMetAsnAspAsnPheAsnIleLeuGl 17

    2592 GACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTG 2641
         |||||||:::::::::|||:::::::||||||   ||||||||||||||||
      17 uThrLeuValThrLeuProLysSerPheIleValLysValProGlySerV 34

    2642 TTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGG 2691
         ||:::|||:::|||   |||||||||   |||||||:::||||||||||||
```



```
4405 .........ACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTT...AC 4442
618 GlnValSerValProValAlaValGluLysThrValLeuSerValGluLy 634

4443 GACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGG 4492
634 sAlaArgGluLeuLysAlaValAspLysGlyLysAlaValValHisAlaL 651

4493 CTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACG 4542
651 ysGluValLysAsnValProValLysThrLeuProArgGlyAlaLeuLys 667

4543 ATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTC 4592
668 IleSerGluAspThrValArgLysGluLeuCysMetPheArgThrCysSe 684

4593 CTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGT 4642
684 rCysGlyValGlnLeuAspValTyrAsnGluAlaThrIleAlaThrArgP 701

4643 TCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTC 4692
701 heSerAsnAlaPheThrPheValAspSerLeuLysGlyArgSerAlaVal 717

4693 TTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGT 4742
718 PhePheSerLysLeuGlyGluGlyTyrThrTyrAsnGlyGlySerHisVa 734

4743 TTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGT 4792
734 lSerSerGlyTrpProArgAlaLeuGluAspIleLeuThrAlaIleLysT 751

4793 ACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGA 4842
751 yrProSerValPheAspHisCysLeuValGlnLysTyrLysMetGlyGly 767

4843 GGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCC 4892
768 GlyValProPheHisAlaAspAspGluGluCysTyrProSerAspAsnPr 784

4893 TATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCA 4942
784 oIleLeuThrValAsnLeuValGlyLysAlaAsnPheSerThrLysCysA 801

4943 GGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTT 4992
801 rgLysGlyGlyLysValMetValIleAsnValAlaSerGlyAspTyrPhe 817

4993 CTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCAT 5042
818 LeuMetProCysGlyPheGlnArgThrHisLeuHisSerValAsnSerIl 834

5043 CGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTG 5092
834 eAspGluGlyArgIleSerLeuThrPheArgAlaThrArgArgValPheG 851

5093 GTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142
851 lyValGlyArgMetLeuGlnLeuAlaGlyGlyValSerAspGluLysSer 867

5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
868 ProGlyValProAsnGlnGlnProGlnSerGlnGlyAlaThrArgThrIl 884

5193 CACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAG 5242
884 eThrProLysSerGlyGlyLysAlaLeuSerGluGlySerGlyArgGluV 901

5243 TCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGG 5292
901 alLysGlyArgSerThrTyrSerIleTrpCysGluGlnAspTyrValArg 917

5293 AAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGA 5342
```

```
918 LysCysGluTrpLeuArgAlaAspAsnProValMetAlaLeuLysProGl 934

5343 CTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATG 5392
934 yTyrThrProMetThrPheGluValValLysAlaGlyThrSerGluAspA 951

5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC 5442
951 laValValGluTyrLeuLysTyrLeuAlaIleGlyIleGlyArgThrTyr 967

5443 AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGT 5492
968 ArgAlaLeuLeuMetAlaArgAsnIleAlaValThrThrAlaGluGlyVa 984

5493 TCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTT 5542
984 lLeuLysValProAsnGlnValTyrGluSerLeuProGlyPheHisValT 1001

5543 ACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGT 5592
1001 yrLysSerGlyThrAspLeuIlePheHisSerThrGlnAspGlyLeuArg 1017

5593 GTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAA 5642
1018 ValArgAspLeuProTyrValPheIleAlaGluLysGlyIlePheIleLy 1034

5643 GGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTAT 5692
1034 sGlyLysAspValAspAlaValValAlaLeuGlyAspAsnLeuSerValC 1051

5693 GCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTG 5742
1051 ysAspAspIleLeuValPheHisAspAlaIleAsnLeuMetGlyAlaLeu 1067

5743 AAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATA 5792
1068 LysValAlaArgCysGlyMetValGlyGluSerPheLysSerPheGluTy 1084

5793 TAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGG 5842
1084 rLysCysTyrAsnAlaProProGlyGlyGlyLysThrThrMetLeuValA 1101

5843 ACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTG 5892
1101 spGluPheValLysSerProAsnSerThrAlaThrIleThrAlaAsnVal 1117

5893 GGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTT 5942
1118 GlySerSerGluAspIleAsnMetAlaValLysLysArgAspProAsnLe 1134

5943 GGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTA 5992
1134 uGluGlyLeuAsnSerAlaThrThrValAsnSerArgValValAsnPheI 1151

5993 TCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATG 6042
1151 leValArgGlyMetTyrLysArgValLeuValAspGluValTyrMetMet 1167

6043 CATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGG 6092
1168 HisGlnGlyLeuLeuGlnLeuGlyValPheAlaThrGlyAlaSerGluGl 1184

6093 CCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGG 6142
1184 yLeuPhePheGlyAspIleAsnGlnIleProPheIleAsnArgGluLysV 1201

6143 TGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTA 6192
1201 alPheArgMetAspCysAlaValPheValProLysLysGluSerValVal 1217

6193 TACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTC 6242
```

```
    1218 TyrThrSerLysSerTyrArgCysProLeuAspValCysTyrLeuLeuSe 1234

    6243 CTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTA 6292
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1234 rSerMetThrValArgGlyThrGluLysCysTyrProGluLysValValS 1251

    6293 GCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGA 6342
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1251 erGlyLysAspLysProValValArgSerLeuSerLysArgProIleGly 1267

    6343 ACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGAC 6392
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1268 ThrThrAspAspValAlaGluIleAsnAlaAspValTyrLeuCysMetTh 1284

    6393 CCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAG 6442
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1284 rGlnLeuGluLysSerAspMetLysArgSerLeuLysGlyLysGlyLysG 1301

    6443 AAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGAT 6492
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1301 luThrProValMetThrValHisGluAlaGlnGlyLysThrPheSerAsp 1317

    6493 GTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACA 6542
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1318 ValValLeuPheArgThrLysLysAlaAspAspSerLeuPheThrLysGl 1334

    6543 ACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATG 6592
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1334 nProHisIleLeuValGlyLeuSerArgHisThrArgSerLeuValTyrA 1351

    6593 CCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGAC 6642
         ||||||||||||||:::|||||||||||||||||||||||||||||||||
    1351 laAlaLeuSerSerGluLeuAspAspLysValGlyThrTyrIleSerAsp 1367

    6643 GCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGC 6692
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1368 AlaSerProGlnSerValSerAspAlaLeuLeuHisThrPheAlaProAl 1384

    6693 TGGTTGCTTTCGAGGTATA 6711
         |||||||||||||||||||
    1384 aGlyCysPheArgGlyIle 1390

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-09-081-320-3

seq_documentation_block:
; Sequence 3, Application US/09081320
; Patent No. 6093544
;  GENERAL INFORMATION:
;    APPLICANT:  Gonsalves, Dennis
;    APPLICANT:  Meng, Baozhong
;    TITLE OF INVENTION:  RUPESTRIS STEM PITTING ASSOCIATED VIRUS
;    TITLE OF INVENTION:  NUCLEIC ACIDS, PROTEINS, AND THEIR USES
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  Clinton Square, P.O. Box 1051
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/081,320
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/047,147
;      FILING DATE:  20-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/069,902
;      FILING DATE:  17-DEC-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1722
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2161 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-081-320-3


alignment_scores:
           Quality:  290.00            Length:    142
             Ratio:  2.736               Gaps:      3
 Percent Similarity:  74.648   Percent Identity: 41.549

alignment_block:
 US-09-301-906-3 x US-09-081-320-3   ..

 Align seg 1/1  to: US-09-081-320-3  from: 1  to: 2161

    4660 TTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGG 4709
         |||   |||:::|||:::|||||||:::|||   ||||||||||||   ||
     733 PheLysAspArgLeuArgGlyArgGlyAlaSerPhePheSerLysProGl 749

    4710 T...GAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGC 4756
         |   :::   ::::::|||||||||||||||||||:::|||   |||||||
     749 yIleSerCysHisSerTyrAsnGlyGlySerHisThrSerLeuGlyTrpP 766

    4757 CTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTC 4806
         ||:::   :::::::::|||||::::::           :::   :::
     766 roLysPheMetAspGlnIleLeuSerSerThrGlyGlyArgAsnTyrTyr 782

    4807 GACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCA 4856
         :::   ||||||:::|||   |||:::   ::::::   :::   :::||
     783 AsnSerCysLeuAlaGlnIleTyrGluGluAsnSerLysLeuAlaLeuHi 799

    4857 CGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCA 4906
         |   ||||||||:::|||||||         :::   :::||||||||||
     799 sLysAspAspGluSerCysTyrGluIleGlyHisLysValLeuThrValA 816

    4907 ATCTCGTGGGGAAGGCAAACTTCTCG...ACTAAGTGCAGGAAG..... 4947
         |||||:::|||:::|||:::|||:::   :::|||   |||:::
     816 snLeuIleGlySerAlaThrPheThrIleSerLysSerArgAsnLeuVal 832

    4948 GGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTAT 4997
         ||||||:::         ::::::::::::   :::::::::|||   ||
     833 GlyGlyAsnHisCysSerLeuThrIleGlyProAsnGluPhePheGluMe 849

    4998 GCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACG 5047
         ||||   |||:::|||   :::::::::|||:::|||::::::
     849 tProArgGlyMetGlnCysAsnTyrPheHisGlyValSerAsnCysThrP 866

    5048 AAGGGCGCATCAGTTTGACGTTCAGG 5073
           ||||||:::|||||||||||||||
     866 roGlyArgValSerLeuThrPheArg 874

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-840-316-1

seq_documentation_block:
; Sequence 1, Application US/08840316
; Patent No. 6054567
;  GENERAL INFORMATION:
;    APPLICANT:  Emerson, Suzanne U., Purcell, Robert H.,
;    APPLICANT:  Tsarev, Sergei. A., and Robinson, Robin A.
```

```
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/840,316
;      FILING DATE:  11-APR-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard W. Bork
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4255
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 AMINO ACID RESIDUES
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  UNKNOWN
;      TOPOLOGY:  UNKNOWN
US-08-840-316-1


alignment_scores:
           Quality:  179.50              Length:    888
             Ratio:  0.481                Gaps:     45
 Percent Similarity:  42.005   Percent Identity: 21.284

alignment_block:
 US-09-301-906-3 x US-08-840-316-1   ..

 Align seg 1/1  to: US-08-840-316-1  from: 1  to: 1693

    4561 GGTAAACAGCTTTGCATGTTTAGAACGTGT.....TCCTGCGG 4598
         |||:::||||||   ::::::      |||          |||:::||
     425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaGl 441

    4599 TGTGCAGCTGGAT.....GTGTACAACGAAGCG..... 4626
         |   :::||||||          |||:::::|||:::
     441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458

    4627 ........ACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
                 |||         :::      :::   |||
     458 isCysArgThrAlaIleArgLysAlaValSerLysPheCysCysPheMet 474

    4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTA 4718
         :::      |||:::::::::::   |||   :::   :::||||||
     475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491

    4719 T.........................ACCTATAATGGT..... 4731
                                   :::|||:::|||
     491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508

    4732 ............................GGTAGCCATGTTTCATCA 4749
                                     ||||||:::|||
     508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524

    4750 GGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAG 4799
         |||         ||||||:::   :::      |||:::      |||::
     525 Gly......ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539

    4800 CGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTAC 4849
         :         :::   :::::::::      ::::::   :::   |||
     539 a.........GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552

    4850 CATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTG 4899
                              :::                |||:::   |||
     553 ....................LysValSerGlnValAspGlyArgIleAsp 562

    4900 ACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGG 4949
               :::|||:::|||:::   :::|||   |||:::
     563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578

    4950 TGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGC 4999
          |||   |||:::      |||
     578 pGlyAlaValLeuGluThrAsn........................... 585

    5000 CTTGCGGTTTTCAAAGGACGCACTTGCAT.........TCAGTAAACTCC 5040
               |||   :::|||   :::|||              |||   ::::::
     586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600

    5041 ATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC..... 5085
         :::      |||   :::||||||||||:::   |||:::
     601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617

    5086 ................GTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
                         :::   |||:::   :::          :::|||
     617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634

    5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCA.............. 5154
           |||||||||   :::::::::||||||
     634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650

    5155 ..............AACCAGCAACCACAG.................... 5169
                       |||::::::   |||
     651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667

    5170 .............AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGG 5206
                        :::|||            ::::::|||   |||
     667 nPheTrpPheHisProGluGlyLeuLeuGlyProPheAlaProPheSerP 684

    5207 GGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCG 5256
         |||   :::   :::      |||::::::          |||:::|||
     684 roGlyHisValTrpGlu......SerAlaAsnProPheCysGlyGluSer 698

    5257 ACATAC......TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
         |||                |||   |||   |||   |||   :::
     699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715

    5301 GTGG................................CTCAGGGCTG 5314
                                             ||||||
     715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732

    5315 ATAATCCAGTGATGGCTCTT..............GAACCTGACTACACC 5349
           :::|||:::   |||                ||||||   :::
     732 laThrProThrProAlaAlaProLeuProProProAlaProAspProSer 748

    5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
         |||   |||:::::::::          |||            ::::::::
     749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764

    5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGT 5449
         :         ::::::          |||      |||   |
     764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781

    5450 TG......................CTTATGGCTAGAAATATTGCCGTC 5475
         ||                      :::::|||   ::::::
     781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797
```

```
5476 ACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT..................... 5508
     :::|||:::      ::::::::::::   |||
 798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814

5509 ................CAAGTTTATGAATCACTACCG.............. 5529
                          |||:::      |||
 814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831

5530 ......................GGCTTTCACGTTTACAAGTCGGGCACA 5556
                           |||      :::|||
 831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847

5557 GATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACC 5606
        |||:::|||::::::            |||::::::      ||
 848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864

5607 G.......................................................... 5607
     |
 864 oLysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881

5608 .....TACGTACTCATAGCTGAAAAAGGTATCTTTACCAAG......GGC 5646
          |||   |||::::::     ||||||:::             |||
 881 laAlaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896

5647 AAAGATGTCGACGCGGTGGTA.........GCTTTGGGCGACAATCTGTT 5687
     :::   ||||||                  |||||||:::|||::
 897 ProSerPheAspAlaTrpGluArgAsnHisArgProGlyAspGluLeuTy 913

5688 CGTA.....................................TGCG 5695
     ::::                                    |||
 913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysP 930

5696 ACGATATACTGGTTTTCCACGAT.....................GCCATT 5724
          :::  :::  :::|||                      ||||||
 930 roThrLeuThrIleThrGluAspValAlaArgThrAlaAsnLeuAlaIle 946

5725 AATTTGATAGGTGCACTGAAAGTCGCTCGATGC................. 5757
     :::|||   :::|||      |||:::|||:::
 947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963

5758 .......GGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCT 5800
           |||:::|||                    :::|||:::
 963 lThrProGlyValVal.....................GlnTyrGlnPheT 973

5801 ATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTC 5850
        :::   ||||||:::||||||:::   ::::::
 973 hrAlaGlyValProGlySerGlyLysSerArgSerIle............. 985

5851 GTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTC 5900
                  :::   |||   :::   :::   |||      ::
 986 ..............ThrGlnAlaAspValAspValValValValProTh 997

5901 TGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTC 5950
     :::::::::::::   |||   :::::::|||                |||:
 997 rArgGluLeuArgAsnAlaTrpArgArgArg..............GlyP 1009

5951 TCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGG 6000
     ::   :::   |||      ::::::      |||:::
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr............. 1021

6001 GGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGG 6050
        :::   :::||||||::::::||||||:::      :::
1022 ...GlnGlyArgArgValValIleAspGluAlaProSerLeuProProHi 1037

6051 CTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTT 6100
      ||||||   |||   :::            |||:::          ::::
1037 sLeuLeuLeuLeuHisMet......GlnArgAlaAlaThrValHisLeuL 1052

6101 TTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGG 6150
     ::||||||   ||||||||||||   |||:::   |||
1052 euGlyAspProAsnGlnIleProAlaIleAspPheGluHis......... 1065

6151 ATGGATTGTGCTGTTTTTGTTCCAAAG...AAGGAAAGCGTTGTATACAC 6197
```

```
                |||   :::||||||      :::   :::::::::   ||
1066 .........AlaGlyLeuValProAlaIleArgProAspLeuAlaProTh 1079

6198 TTCT.........AAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGT 6238
     ||||            :::::|||||||||   ||||||||   |||
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095

6239 TGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTC 6288
                    :::||||||         :::||||||
1096 ..............IleArgGly.........AlaTyrProMetIleGln 1104

6289 GTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG......TCCAAAAG 6332
     ::::::::::::         |||:::||||||||      ::::::
1105 ThrThrSerArg.........ValLeuArgSerLeuPheTrpGlyGluPr 1118

6333 GCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACT 6382
         :::|||                                      |
1118 oAlaValGlyGlnLys................................L 1124

6383 TGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGA 6432
     ||   :::||||||      |||:::            |||
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer......... 1137

6433 AAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAAC 6482
                         :::||||||||||||||||||||||   ||
1138 .....................ValThrValHisGluAlaGlnGlyAlaTh 1147

6483 ATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTAT 6532
     |:::::::::::::::::::   |||      |||         :::
1147 rTyrThrGluThrThrIleIleAlaThrAlaAspAlaArgGlyLeuIleG 1164

6533 TCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCA 6582
       :::::::::   |||   :::|||:::|||:::|||||||||:::::
1164 lnSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180

6583 CTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATA 6632
       |||                :::   |||   :::   ||||||
1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly..... 1194

6633 TATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGT 6682
                                  :::|||||||||:::::::::::|
1195 .........................IleSerAspAlaIleValAsnAsnP 1203

6683 TCGCCCCGGCTGGT 6696
     ||      ||||||
1203 hePheLeuAlaGly 1207

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-08849A-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9308849A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08849A
;      FILING DATE:  17-SEP-1993
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US07/947,263
;      FILING DATE:  18-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William S. Feiler
;      REGISTRATION NUMBER:  26,728
;      REFERENCE/DOCKET NUMBER:  2026-4032 PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 AMINO ACID RESIDUES
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  UNKNOWN
;      TOPOLOGY:  UNKNOWN
PCT-US93-08849A-1


alignment_scores:
           Quality:  179.50              Length:    888
             Ratio:  0.481                Gaps:     45
 Percent Similarity:  42.005   Percent Identity: 21.284

alignment_block:
 US-09-301-906-3 x PCT-US93-08849A-1   ..

 Align seg 1/1  to: PCT-US93-08849A-1  from: 1  to: 1693

    4561 GGTAAACAGCTTTGCATGTTTAGAACGTGT...........TCCTGCGG 4598
         |||:::||||||   ::::::     |||          |||:::||
     425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaGl 441

    4599 TGTGCAGCTGGAT...........GTGTACAACGAAGCG.......... 4626
         |   :::||||||           |||:::::|||:::
     441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458

    4627 ........ACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
                 |||         :::      :::   |||
     458 isCysArgThrAlaIleArgLysAlaValSerLysPheCysCysPheMet 474

    4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTA 4718
         :::      |||:::::::::::::   |||   :::   :::||||||
     475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491

    4719 T..........................ACCTATAATGGT.......... 4731
                                     :::|||:::|||
     491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508

    4732 ................................GGTAGCCATGTTTCATCA 4749
                                        ||||||:::|||
     508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524

    4750 GGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAG 4799
         |||         ||||||:::   :::      |||:::      |||::
     525 Gly......ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539

    4800 CGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTAC 4849
         :         :::   :::::::::      ::::::   :::   |||
     539 a.........GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552

    4850 CATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTG 4899
                            :::            |||:::   |||
     553 ...................LysValSerGlnValAspGlyArgIleAsp 562

    4900 ACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGG 4949
               :::|||:::|||:::   :::|||   |||:::
     563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578

    4950 TGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGC 4999
          |||   |||:::      |||
     578 pGlyAlaValLeuGluThrAsn........................... 585

    5000 CTTGCGGTTTTCAAAGGACGCACTTGCAT.........TCAGTAAACTCC 5040

          |||   :::|||   :::|||           |||   ::::::
     586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600

    5041 ATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC..... 5085
         :::       |||   :::||||||||||:::   |||:::
     601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617

    5086 ..............GTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
                         :::   |||:::   :::          :::|||
     617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634

    5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCA.............. 5154
           |||||||||   :::::::::||||||
     634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650

    5155 ..............AACCAGCAACCACAG................... 5169
                        |||:::::::   |||
     651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667

    5170 ............AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGG 5206
                       :::|||            ::::::|||   |||
     667 nPheTrpPheHisProGluGlyLeuLeuGlyProPheAlaProPheSerP 684

    5207 GGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCG 5256
           |||   :::   :::      |||::::::         |||:::|||
     684 roGlyHisValTrpGlu......SerAlaAsnProPheCysGlyGluSer 698

    5257 ACATAC......TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
         |||                |||   |||   |||   |||   :::
     699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715

    5301 GTGG.................................CTCAGGGCTG 5314
                                                ||||||
     715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732

    5315 ATAATCCAGTGATGGCTCTT..............GAACCTGACTACACC 5349
           :::|||:::   |||                 ||||||   :::
     732 laThrProThrProAlaAlaProLeuProProProAlaProAspProSer 748

    5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
         |||   |||::::::::::         |||              ::::::::
     749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764

    5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGT 5449
         :              ::::::            |||      |||   |
     764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781

    5450 TG.....................CTTATGGCTAGAAATATTGCCGTC 5475
         ||                     :::::|||   ::::::
     781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797

    5476 ACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT................. 5508
         :::|||:::   ::::::::::::   |||
     798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814

    5509 ..............CAAGTTTATGAATCACTACCG.............. 5529
                              |||:::      |||
     814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831

    5530 ....................GGCTTTCACGTTTACAAGTCGGGCACA 5556
                              |||      :::|||
     831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847

    5557 GATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACC 5606
           |||:::|||::::::         |||::::::      ||
     848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864

    5607 G..................................................... 5607
         |
     864 oLysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881

    5608 .....TACGTACTCATAGCTGAAAAAGGTATCTTTACCAAG......GGC 5646
               |||   |||::::::      ||||||:::            |||
```

```
 881 laAlaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896

5647 AAAGATGTCGACGCGGTGGTA.........GCTTTGGGCGACAATCTGTT 5687
       :::   ||||||                  ||||||:::|||::
 897 ProSerPheAspAlaTrpGluArgAsnHisArgProGlyAspGluLeuTy 913

5688 CGTA.........................................TGCG 5695
     ::::                                        |||
 913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysP 930

5696 ACGATATACTGGTTTTCCACGAT.....................GCCATT 5724
          :::   :::   :::|||                    ||||||
 930 roThrLeuThrIleThrGluAspValAlaArgThrAlaAsnLeuAlaIle 946

5725 AATTTGATAGGTGCACTGAAAGTCGCTCGATGC................. 5757
     :::|||   :::|||      |||:::|||:::
 947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963

5758 .......GGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCT 5800
           |||:::|||                  :::|||:::
 963 lThrProGlyValVal.....................GlnTyrGlnPheT 973

5801 ATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTC 5850
          :::   ||||||:::|||||||:::   ::::::
 973 hrAlaGlyValProGlySerGlyLysSerArgSerIle............ 985

5851 GTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTC 5900
                    :::   |||   :::   :::   |||      ::
 986 ..............ThrGlnAlaAspValAspValValValValProTh 997

5901 TGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTC 5950
     ::::::::::::   |||   :::::::|||               |||:
 997 rArgGluLeuArgAsnAlaTrpArgArgArg..............GlyP 1009

5951 TCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGG 6000
     ::   :::   |||      ::::::      |||:::
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr............ 1021

6001 GGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGG 6050
        :::   :::||||||::::::||||||:::      :::
1022 ...GlnGlyArgArgValValIleAspGluAlaProSerLeuProProHi 1037

6051 CTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTT 6100
      ||||||   |||   :::           |||:::        ::::
1037 sLeuLeuLeuLeuHisMet......GlnArgAlaAlaThrValHisLeuL 1052

6101 TTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGG 6150
     ::||||||   ||||||||||||   |||:::   |||
1052 euGlyAspProAsnGlnIleProAlaIleAspPheGluHis......... 1065

6151 ATGGATTGTGCTGTTTTTGTTCCAAAG...AAGGAAAGCGTTGTATACAC 6197
               |||   :::||||||      :::   :::::::::   ||
1066 .........AlaGlyLeuValProAlaIleArgProAspLeuAlaProTh 1079

6198 TTCT.........AAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGT 6238
     ||||              ::::::||||||||||   |||||||||   |||
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095

6239 TGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTC 6288
                   :::||||||         :::||||||
1096 ..............IleArgGly.........AlaTyrProMetIleGln 1104

6289 GTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG......TCCAAAAG 6332
     ::::::::::::         |||:::|||||||||      ::::::
1105 ThrThrSerArg.........ValLeuArgSerLeuPheTrpGlyGluPr 1118

6333 GCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACT 6382
          :::|||                                     |
1118 oAlaValGlyGlnLys................................L 1124

6383 TGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGA 6432
     ||   :::||||||      |||:::            |||
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer......... 1137

     6433 AAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAAC 6482
                                  :::|||||||||||||||||||||   ||
     1138 .....................ValThrValHisGluAlaGlnGlyAlaTh 1147

     6483 ATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTAT 6532
          |:::::::::::::::::::   |||      |||         :::
     1147 rTyrThrGluThrThrIleIleAlaThrAlaAspAlaArgGlyLeuIleG 1164

     6533 TCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCA 6582
            :::::::::   |||   :::|||:::|||:::||||||||||::::::
     1164 lnSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180

     6583 CTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATA 6632
             |||                :::   |||   :::   ||||||
     1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly..... 1194

     6633 TATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGT 6682
                                      :::|||||||||:::::::::::::|
     1195 .........................IleSerAspAlaIleValAsnAsnP 1203

     6683 TCGCCCCGGCTGGT 6696
          ||       ||||||
     1203 hePheLeuAlaGly 1207
```

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US93-08849-1

```
seq_documentation_block:
; Sequence 1, Application PC/TUS9308849
;  GENERAL INFORMATION:
;    APPLICANT:  Tsarev, Sergei A., Emerson,
;    APPLICANT:  Suzanne U., Purcell, Robert H.
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08849
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/947,263
;      FILING DATE:  18-SEP-1992
;      NAME:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bork, Richard, W.
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 amino acid residues
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
PCT-US93-08849-1


alignment_scores:
```

```
          Quality:  179.50            Length:    888
            Ratio:  0.481               Gaps:     45
 Percent Similarity:  42.005   Percent Identity: 21.284

alignment_block:
 US-09-301-906-3 x PCT-US93-08849-1   ..

 Align seg 1/1  to: PCT-US93-08849-1  from: 1  to: 1693

   4561 GGTAAACAGCTTTGCATGTTTAGAACGTGT...........TCCTGCGG 4598
        |||:::|||||   ::::::      |||           |||:::||
    425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaGl 441

   4599 TGTGCAGCTGGAT...........GTGTACAACGAAGCG......... 4626
        |   :::||||||           |||:::::|||:::
    441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458

   4627 ........ACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
                |||         :::      :::   |||
    458 isCysArgThrAlaIleArgLysAlaValSerLysPheCysCysPheMet 474

   4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTA 4718
        :::      |||:::::::::::::   |||   :::   :::||||||
    475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyVa 491

   4719 T.........................ACCTATAATGGT......... 4731
                                  :::|||:::|||
    491 lValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508

   4732 ...............................GGTAGCCATGTTTCATCA 4749
                                       ||||||:::|||
    508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524

   4750 GGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAG 4799
        |||         ||||||:::   :::      |||:::      |||::
    525 Gly......ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539

   4800 CGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTAC 4849
        :         :::   :::::::::      ::::::   :::   |||
    539 a.........GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552

   4850 CATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTG 4899
                         :::           |||:::   |||
    553 ...................LysValSerGlnValAspGlyArgIleAsp 562

   4900 ACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGG 4949
              :::|||:::|||:::   :::|||   |||:::
    563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578

   4950 TGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGC 4999
         |||   |||::      |||
    578 pGlyAlaValLeuGluThrAsn........................... 585

   5000 CTTGCGGTTTTCAAAGGACGCACTTGCAT.........TCAGTAAACTCC 5040
             |||   :::|||  .:::|||            |||   ::::::
    586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600

   5041 ATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC..... 5085
        :::      |||   :::||||||||::   |||:::
    601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617

   5086 ................GTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
                        :::   |||:::   :::          :::|||
    617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634

   5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCA............... 5154
          |||||||||   :::::::::||||||
    634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650

   5155 ..............AACCAGCAACCACAG.................... 5169
                       |||:::::   |||
    651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgLeuSerLeuThrGlyAs 667

   5170 ............AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGG 5206
                       :::|||          :::::::|||   |||
    667 nPheTrpPheHisProGluGlyLeuLeuGlyProPheAlaProPheSerP 684

   5207 GGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCG 5256
          |||   :::   :::      |||::::::          |||:::|||
    684 roGlyHisValTrpGlu......SerAlaAsnProPheCysGlyGluSer 698

   5257 ACATAC......TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
        |||               |||   |||   |||   |||   :::
    699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValProSerProAl 715

   5301 GTGG................................CTCAGGGCTG 5314
                                            ||||||
    715 aGlnProAspLeuGlyPheThrSerGluProSerIleProSerArgAlaA 732

   5315 ATAATCCAGTGATGGCTCTT..............GAACCTGACTACACC 5349
          :::|||:::   |||                 ||||||   :::
    732 laThrProThrProAlaAlaProLeuProProProAlaProAspProSer 748

   5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
        |||   |||:::::::::         |||            ::::::::
    749 Pro...ThrLeuSerAlaProAlaArgGlyGluProAlaProGlyAlaTh 764

   5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGT 5449
        :         ::::::          |||      |||   |
    764 rAlaArgAlaProAlaIleThrHisGlnThrAlaArgHisArgArgLeuL 781

   5450 TG.....................CTTATGGCTAGAAATATTGCCGTC 5475
        ||                    ::::::|||   ::::::
    781 euPheThrTyrProAspGlySerLysValPheAlaGlySerLeuPheGlu 797

   5476 ACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT................. 5508
        :::|||:::      ::::::::::::   |||
    798 SerThrCysThrTrpLeuValAsnAlaSerAsnValAspHisArgProGl 814

   5509 ................CAAGTTTATGAATCACTACCG............. 5529
                           |||:::      |||
    814 yGlyGlyLeuCysHisAlaPheTyrGlnArgTyrProAlaSerPheAspA 831

   5530 ......................GGCTTTCACGTTTACAAGTCGGGCACA 5556
                             |||      ::::||
    831 laAlaSerPheValMetArgAspGlyAlaAlaAlaTyrThrLeuThrPro 847

   5557 GATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACC 5606
              |||:::|||::::::          |||::::::      ||
    848 ArgProIleIleHisAlaValAlaProAspTyrArgLeuGluHisAsnPr 864

   5607 G..................................................... 5607
        |
    864 oLysArgLeuGluAlaAlaTyrArgGluThrCysSerArgLeuGlyThrA 881

   5608 .....TACGTACTCATAGCTGAAAAAGGTATCTTTACCAAG......GGC 5646
             |||   |||::::::      ||||||:::            |||
    881 laAlaTyrProLeuLeuGly...ThrGlyIleTyrGlnValProIleGly 896

   5647 AAAGATGTCGACGCGGTGGTA.........GCTTTGGGCGACAATCTGTT 5687
          :::   ||||||                ||||||:::|||::
    897 ProSerPheAspAlaTrpGluArgAsnHisArgProGlyAspGluLeuTy 913

   5688 CGTA.....................................TGCG 5695
        ::::                                     |||
    913 rLeuProGluLeuAlaAlaArgTrpPheGluAlaAsnArgProThrCysP 930

   5696 ACGATATACTGGTTTTCCACGAT.....................GCCATT 5724
             :::   :::   :::|||                    ||||||
    930 roThrLeuThrIleThrGluAspValAlaArgThrAlaAsnLeuAlaIle 946

   5725 AATTTGATAGGTGCACTGAAAGTCGCTCGATGC................. 5757
        :::|||   :::|||      |||:::|||:::
    947 GluLeuAspSerAlaThrAspValGlyArgAlaCysAlaGlyCysArgVa 963

   5758 .......GGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCT 5800
```

```
          |||:::|||                    :::|||:::
 963 lThrProGlyValVal.....................GlnTyrGlnPheT 973

5801 ATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTC 5850
       :::   ||||||:::||||||:::   ::::::
 973 hrAlaGlyValProGlySerGlyLysSerArgSerIle............ 985

5851 GTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTC 5900
                   :::   |||   :::   :::   |||      ::
 986 ..............ThrGlnAlaAspValAspValValValValProTh 997

5901 TGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTC 5950
     :::::::::::::   |||   :::::::|||                |||:
 997 rArgGluLeuArgAsnAlaTrpArgArgArg..............GlyP 1009

5951 TCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGG 6000
     ::   :::   |||      ::::::      |||:::
1009 heAlaAlaPheThrProHisThrAlaAlaArgValThr............ 1021

6001 GGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGG 6050
        :::   :::|||||||::::::|||||||:::      :::
1022 ...GlnGlyArgArgValValIleAspGluAlaProSerLeuProProHi 1037

6051 CTTACTACAACTAGGCGTCTTCGCAACCGGCCGCGTCGGAAGGCCTCTTTT 6100
      ||||||   |||   :::            |||:::      ::::
1037 sLeuLeuLeuLeuHisMet......GlnArgAlaAlaThrValHisLeuL 1052

6101 TTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGG 6150
     ::||||||   ||||||||||||   |||:::   |||
1052 euGlyAspProAsnGlnIleProAlaIleAspPheGluHis......... 1065

6151 ATGGATTGTGCTGTTTTTGTTCCAAAG...AAGGAAAGCGTTGTATACAC 6197
               |||   :::||||||      :::   :::::::::   ||
1066 .........AlaGlyLeuValProAlaIleArgProAspLeuAlaProTh 1079

6198 TTCT.........AAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGT 6238
     ||||           ::::::|||||||||   |||||||||   |||
1079 rSerTrpTrpHisValThrHisArgCysProAlaAspValCysGluLeu. 1095

6239 TGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTC 6288
                  :::||||||          :::||||||
1096 ..............IleArgGly.........AlaTyrProMetIleGln 1104

6289 GTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG......TCCAAAAG 6332
     ::::::::::::          |||:::|||||||||      ::::::
1105 ThrThrSerArg.........ValLeuArgSerLeuPheTrpGlyGluPr 1118

6333 GCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACT 6382
        :::|||                                      |
1118 oAlaValGlyGlnLys................................L 1124

6383 TGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGA 6432
     ||   :::||||||      |||:::            |||
1124 euValPheThrGlnAlaAlaLysAlaAlaAsnProGlySer......... 1137

6433 AAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAAC 6482
                        :::|||||||||||||||||||||   ||
1138 .....................ValThrValHisGluAlaGlnGlyAlaTh 1147

6483 ATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTAT 6532
     |:::::::::::::::::::   |||      |||        :::
1147 rTyrThrGluThrThrIleIleAlaThrAlaAspAlaArgGlyLeuIleG 1164

6533 TCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCA 6582
       :::::::::   |||   :::|||:::|||:::||||||||||::::::
1164 lnSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThrGluLys 1180

6583 CTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATA 6632
        |||               :::   |||   :::   ||||||
1181 CysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGly..... 1194

6633 TATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGT 6682
                              :::|||||||||:::::::::::::|
```

```
     1195 .......................IleSerAspAlaIleValAsnAsnP 1203

     6683 TCGCCCCGGCTGGT 6696
          ||       ||||||
     1203 hePheLeuAlaGly 1207

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-325-267A-2

seq_documentation_block:
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-2


alignment_scores:
            Quality:  178.00              Length:   1508
              Ratio:  0.287                 Gaps:     67
 Percent Similarity:  41.180   Percent Identity: 19.098

alignment_block:
 US-09-301-906-3/rev x US-08-325-267A-2   ..

 Align seg 1/1  to: US-08-325-267A-2  from: 1  to: 1537

    5504 GGTACTTTCAGAACACCTTCGGCGGTA.........GTGACGGCAATATT 5464
         |||   :::   ||||||:::   |||          :::|||:::   ||
     128 GlyPheTyrThrThrProThrAsnValThrLeuGluMetThrGlyTyrPh 144
```

```
5463 TCTAGCCATAAGCAACGCCCTGTATGTCCTCTCAATGCCTATAGCCAGAT 5414
144 eLeuProProGlnThrGlySerTyrThr..... 153

5413 ACTTCAAGTACTCCACGACGGCATCTTCAGAGGTC..... 5379
154 ..PheLysPheAlaThrValAspAspSerAlaIleLeuSerValGlyGly 169

5378 .....CCGGTTTTAACCACTTC 5362
170 AlaThrAlaPheAsnCysCysAlaGlnGlnGlnProProIleThrSerTh 186

5361 AAATGTCATTGGGGTGTAGTCAGGTTCAAGAGCCATCACTGGATTATCAG 5312
186 rAsnPhe.....ThrIleAspGlyIleLysP 195

5311 CCCTGAGCCACTCA..... 5298
195 roTrpGlyGlySerLeuProProAsnIleGluGlyThrValTyrMetTyr 211

5297 .....CACTTCCTAACGTAATCTTGTTCGCACCATATCGAGTA 5260
212 AlaGlyTyrTyrTyrProMetLysValValTyrSerAsnAlaValSerTr 228

5259 TGTCGACCTCCCCTTGACTTCCCTACCACTTCCCTCAGATAGAGCCTTGC 5210
228 pGlyThrLeuProIle...SerValThrLeuProAspGlyThrThrValS 244

5209 CCCCCGATTTTGGTGTGATTGTTCTGGTAGCACCTTGGCTCTGTGGTTGC 5160
244 erAspAspPheGlu..... 248

5159 TGGTTTGGAACACCTGGTGACTTCTCATCCGACACGCCGCCGGCTAACTG 5110
249 .....GlyTyrValTyrSerPheAspAspAspLeuSerGlnSerAsnCy 263

5109 CAACATCCTGCCTACACCAAAGACGCGCCGAGTTGCCCTGAACGTCAAAC 5060
263 sThrVal...ProAspProSerAsnTyrAlaValSerThrThrThrThrT 279

5059 TGATGCGCCCTTCGTCGATGGAGTTTACTGAATGCAAGTGCGTCCTTTGA 5010
279 hrThrGluProTrpThrGlyThrPheThrSerThr..... 290

5009 AAACCGCAAGGCATAAGAAAATAGTCACCCGAAGCTACGTTTATGACCAT 4960
291 .....SerThrGluMetThrThrValThrGl 299

4959 GACCTTA...CCACCCTTCCTGCACTTAGTCGAGAAGTTTGCCTTCCCCA 4913
299 yThrAsnGlyValProThrAspGluThrValIleValIleArgThrProT 316

4912 CGAGATTGACCGTCAAGATAGGGTTATCTGATGGATAGCACTCCTCGTCA 4863
316 hrThrAlaSerThrIleIle.....ThrThrThr 325

4862 TCAGCGTGGAATGGTACGCCTCCACCCATCTTGTACTTCTGCACTAAACA 4813
326 GluProTrpAsnSerThr..... 331

4812 GTGGTCGAAGACGCTTGGGTACTTAATTGCCGTTAAGATATCCTCTAGGG 4763
332 .PheThrSerThrSerThrGluLeuThrThrValThrGlyThrAsnGlyV 348

4762 CACGAGGCCACCCTGATGAAACATGGCTACCACCATTATAGGTATACCCC 4713
348 alArg.....ThrAspGluThrIleIle.....ValIleArg 358

4712 TCACCCAGCTTTGAGAAAAAGACCGCACTCCTCCCTTTCAAGTTATCGAC 4663
359 ThrProThr.....ThrAlaThrThrAlaIleThrThrThrGl 371

4662 AAAGGTAAATGCGTTTGAGAACCTGGTGGCGATAGTCGCTTCGTTGTACA 4613
371 uProTrpAsn.....SerThrPheT 378

4612 CATCCAGCTGCACACCGCAGGAACACGTTCTAAACATGCAAAGCTGTTTA 4563
378 hrSerThrSerThrGluLeuThrThrValThrGlyThrAsnGly...Leu 393

4562 CCGGCCTTGCCTTCATTGATCGTCAAACTCCTCTGTTTGGGCTGC..... 4518
394 ProThrAspGluThrIleIleValIleArgThrProThrThrAlaThrTh 410

4517 .....TCGACTTGTA 4508
410 rAlaMetThrThrThrGlnProTrpAsnAspThrPheThrSerThrSerT 427

4507 CCTGTTTTTTCTCAGCCACATGAGCGACGGCCTTGCCC..... 4470
427 hrGluMetThrThrValThrGlyThrAsnGlyLeuProThrAspGluThr 443

4469 TTGTCAATTATCTTCACACCACGAGTCGTAACACTTGGGGCG..... 4428
444 IleIleValIleArgThrProThrThrAlaThrThrAlaMetThrThrTh 460

4427 .....GGTTTCTCATCGACG...ATTGGCGTGGCGCCCG 4397
460 rGlnProTrpAspAspThrPheThrSerThrSerThrGluMetThrThrV 477

4396 AAACAGTCGCCAATGGT..... 4380
477 alThrGlyThrAsnGlyLeuProThrAspGluThrIleIleValIleArg 493

4379 ......ACTTCTTGTTTCACTTCTTTCTGAGCATGTACCTCGGACTGGGT 4336
494 ThrProThrThrAlaThrThrAlaMet...ThrThrThrGlnProTrpAs 509

4335 GACCTTTGGAATATCAGCCTGTGGCGAACTCACAAGACTGGACTGTACTG 4286
509 nAspThrPheThrSerThrSerThrGluMetThrThrValThrGlyThrA 526

4285 GCTCCCTGTCGCCTTCGTTGACCGCGGTTTGTGCTGTGTTACCATCTGCA 4236
526 snGlyLeuProThrAspGluThrIleIleValIleArgThrProThrThr 542

4235 GCTCTT.....CCCCTTTCTTG 4219
543 AlaThrThrAlaMetThrThrThrGlnProTrpAsnAspThrPheThrSe 559

4218 TGTCGGCGCCTCACCGACGACGCTTTCTGGGAGCAACCCTTTGCCGTCC. 4170
559 rThrSerThrGluMetThrThrValThrGlyThrAsnGlyLeuProThrA 576

4169 .....ACCTCAACCTCACGAGCCGTT 4149
576 spGluThrIleIleValIleArgThrProThrThrAlaThrThrAlaIle 592

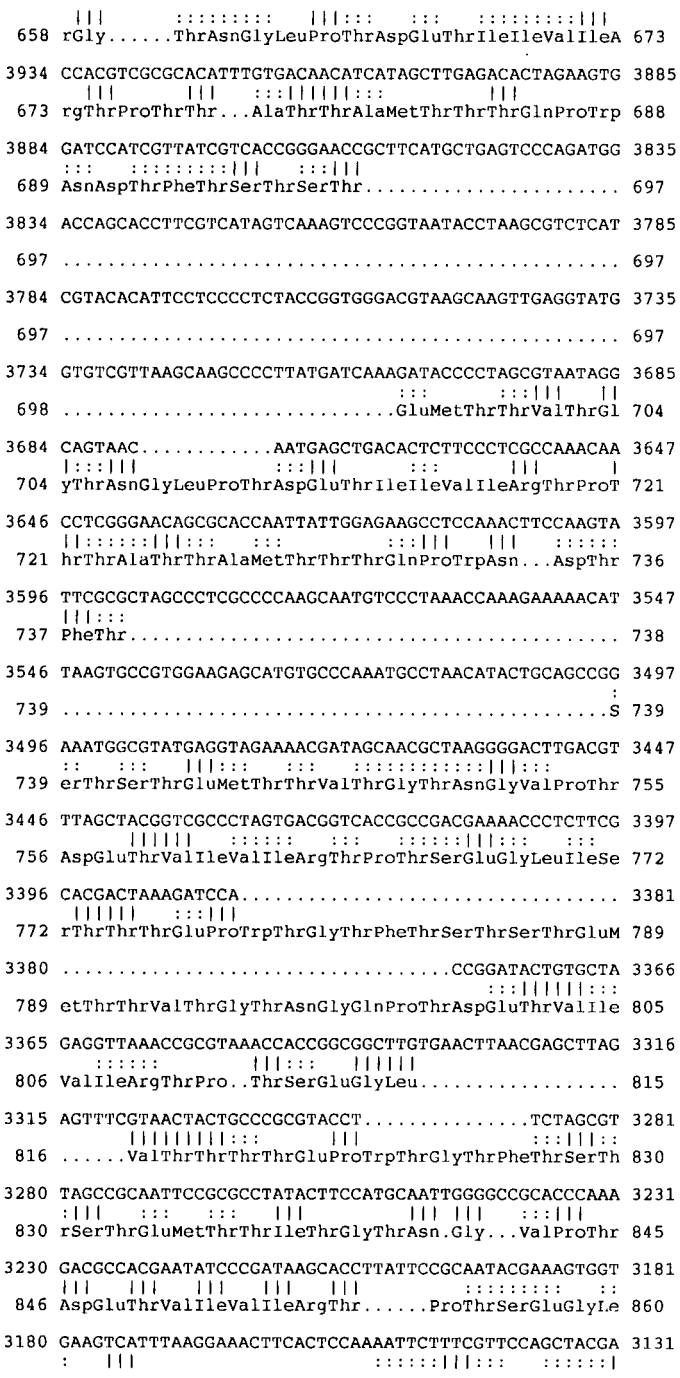
4148 TCCCCACTAGACTGTTGGTGAAAGACATCGTCGACACCTTTTTTCACACT 4099
593 ThrThrThrGluProTrpAsnSerThrPheThrSerThrSerThrGluLe 609

4098 TTCAACAAAAGCCATAGCGTCAGCGGTACCAGAACCCGAC..... 4059
609 uThrThrValThrGlyThrAsnGlyLeuProThrAspGluThrIleIleV 626

4058 .....GGATCACCTTGTCTTTCCGCCACAGCTGCA..... 4029
626 alIleArgThrProThrThrAlaThrThrAlaMetThrThrThrGlnPro 642

4028 ......GACGGGGCAACCTCTTCCTCGACATAAGTGTATTGCACGCTTCG 3985
643 TrpAsnAspThrPheThrSerThrSerThr...GluMetThrThrValTh 658

3984 AGGTGAATGGGTCTCTTCTTCACCAGTAACTTCCCCGTTGGTGCTAATCC 3935
```


658 rGly.....ThrAsnGlyLeuProThrAspGluThrIleIleValIleA 673
3934 CCACGTCGCGCACATTTGTGACAACATCATAGCTTGAGACACTAGAAGTG 3885
673 rgThrProThrThr...AlaThrThrAlaMetThrThrThrGlnProTrp 688
3884 GATCCATCGTTATCGTCACCGGGAACCGCTTCATGCTGAGTCCCAGATGG 3835
689 AsnAspThrPheThrSerThrSerThr..... 697
3834 ACCAGCACCTTCGTCATAGTCAAAGTCCCGGTAATACCTAAGCGTCTCAT 3785
697 697
3784 CGTACACATTCCTCCCCTCTACCGGTGGGACGTAAGCAAGTTGAGGTATG 3735
697 697
3734 GTGTCGTTAAGCAAGCCCCTTATGATCAAAGATACCCCTAGCGTAATAGG 3685
698 GluMetThrThrValThrGl 704
3684 CAGTAAC.....AATGAGCTGACACTCTTCCCTCGCCAAACAA 3647
704 yThrAsnGlyLeuProThrAspGluThrIleIleValIleArgThrProT 721
3646 CCTCGGGAACAGCGCACCAATTATTGGAGAAGCCTCCAAACTTCCAAGTA 3597
721 hrThrAlaThrThrAlaMetThrThrThrGlnProTrpAsn...AspThr 736
3596 TTCGCGCTAGCCCTCGCCCCAAGCAATGTCCCTAAACCAAAGAAAAACAT 3547
737 PheThr..... 738
3546 TAAGTGCCGTGGAAGAGCATGTGCCCAAATGCCTAACATACTGCAGCCGG 3497
739 S 739
3496 AAATGGCGTATGAGGTAGAAAACGATAGCAACGCTAAGGGGACTTGACGT 3447
739 erThrSerThrGluMetThrThrValThrGlyThrAsnGlyValProThr 755
3446 TTAGCTACGGTCGCCCTAGTGACGGTCACCGCCGACGAAAACCCTCTTCG 3397
756 AspGluThrValIleValIleArgThrProThrSerGluGlyLeuIleSe 772
3396 CACGACTAAAGATCCA..... 3381
772 rThrThrThrGluProTrpThrGlyThrPheThrSerThrSerThrGluM 789
3380 CCGGATACTGTGCTA 3366
789 etThrThrValThrGlyThrAsnGlyGlnProThrAspGluThrValIle 805
3365 GAGGTTAAACCGCGTAAACCACCGGCGGCTTGTGAACTTAACGAGCTTAG 3316
806 ValIleArgThrPro..ThrSerGluGlyLeu..... 815
3315 AGTTTCGTAACTACTGCCCGCGTACCT.....TCTAGCGT 3281
816 ValThrThrThrThrGluProTrpThrGlyThrPheThrSerTh 830
3280 TAGCCGCAATTCCGCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAA 3231
830 rSerThrGluMetThrThrIleThrGlyThrAsn.Gly...ValProThr 845
3230 GACGCCACGAATATCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGT 3181
846 AspGluThrValIleValIleArgThr.....ProThrSerGluGlyLe 860
3180 GAAGTCATTTAAGGAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGA 3131
860 uIleSerThrThrThrGluProTrpThrGlyThrPheThrSerThrSerT 877
3130 CCTTGCTCACAGCTACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTT 3081
877 hrGluMetThrThrIleThrGlyThrAsnGlyGlnProThrAspGluThr 893
3080 ATAGCATTCTCTGTGCGCACCAGAGCATCAAGC.....CTTTCATCTTC 3037
894 Val...IleValIleArgThrProThrSerGluGlyLeuIleSerThrTh 909
3036 AGTTAAACCG.....CGGTCGTAGAATAGTGAACCTGTGGAAAAGAACG 2993
909 rThrGluProTrpThrGlyThrPheThrSerThrSerThrGluMetThrT 926
2992 ACGTAATGCCAGAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAACCC 2943
926 hrVal.....ThrGlyThrAsnGly 932
2942 TTAGAAACACTCGAAACCACGGTCGCTTTCGAACCGCCAAGAAGACCTGC 2893
933 ValProThrAspGluThrValIleValIleArgThrProThrSerGluGl 949
2892 CCTCGTACAGTGGACGTTAGAAACGTTCCCCATTATGAACGCATATACAG 2843
949 yLeuIleSerThrThrThrGlu.....ProTrpThrGlyThrPheThrS 964
2842 AATCTTCA.....GTAATTGGC...AGGGGTTTTCCATCT 2811
964 erThrSerThrGluValThrThrIleThrGlyThrAsnGlyGlnProThr 980
2810 TCTGTCTTCATTGCCTTCATGACCTTAATTGAATCTTCCACGATAGCCCT 2761
981 AspGluThrValIleVa 986
2760 AGAGAAGACGCGCAAGCGACTATTCTTCAGTTTCCTACCGAAGGTCTCCT 2711
986 lIleArgThr..... 989
2710 TCGAAATATCAAAGGCACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAA 2661
990 ProThrSerGluGlyLeu..... 995
2660 GTGCAAATGGTAACAACAACAGACCCAGGTACTTTACGGATGAACGCTCT 2611
996 IleSerThrThrThrGluProTrpThrGlyThrPheThrSerTh 1010
2610 GGGCATAGACATAAGCGTCTCCAAAACGTTGAAGTTATCGTGCACCATTG 2561
1010 rSerThrGluMetThrThr.....ValThrGlyThrAsnG 1022
2560 CCTTAGCATAGACGGAGACTGCGTTCGAAAAACAATCGCCTACGGTCTGA 2511
1022 lyGlnProThrAspGluThrValIleValIleArgThrProThr..... 1036
2510 AAGAAGAGCTTGAAGACCCCGGCGAGACTCCCCTCACTGGCTTTTATAAA 2461
1037 SerGluGlyLeu..... 1040
2460 GTTATAGGAGCGAATACTCTTCGCCCTATCCTTAATCGCCTGAGCCATAA 2411
1041 ValT 1042
2410 CAGCGACAACTAAACCCCACATA.....TCATCAGGAGAA 2376
1042 hrThrThrThrGluProTrpThrGlyThrPheThrSerThrSerThrGlu 1058
2375 ATATCAACCTTGCTGTGGACAACGCGACTCCCTGTAATGACATGGGTCTT 2326
1059 MetSerThrValThrGlyThrAsnGlyLeuProThrAspGluThrValIl 1075
2325 TTGCGACCTGACCGCCCCAACGGCATACTCAAACGTTCTATTCTCGAATG 2276
1075 eValValLysThrProThrThrAlaIleSerSerSerLeuSerSer.... 1090

```
  2275 TCCCTATAGTGTTGAGCGCGTAGGAATAGATCCTGTCGACAAAGTCCGAA 2226
                    |||:::         |||   |||:::   ::::::
  1091 ..............SerSerSerGlyGlnIleThrSerSerIleThrSer 1102

  2225 TCGAGGACTATTGTCTTAAAGGACACACTAGAACCAGCTACA...ACAGG 2179
       ||||||   |||:::    .         :::   |||::::::   |||::
  1103 SerArgProIleIle......ThrProPheTyrProSerAsnGlyThrSe 1117

  2178 TATGAACACCAGCGATTTGCCCACGAACGACGGGACCAACGACCTATACG 2129
       ::::   :::|||::::::   :::::::::   |||:::::::|||:::
  1117 rValIleSerSerSerValIleSerSerSerValThrSerSerLeuPheT 1134

  2128 TAACCTCAGTTCCAGGTGACTTCTGAGCCCTACACATAGTGAGATGATGG 2079
       :::::|||
  1134 hrSerSer..................................... 1136

  2078 TAACCACAACGGTATCCCTCATACTCTAGCTTAAACACGTTCCCGGACGA 2029
                    |||      ||||||
  1137 ..............ProValIleSerSer..................... 1141

  2028 AGTGCGTACATAAGAAAAGGTGAAAAACCCGCTTACGTTAGAGAAGGAAT 1979

  1141 ................................................. 1141

  1978 GTTCATACATCTCACCACACTGACCAACATTGTAGGCCAGGTAATCACCT 1929
         |||   ||||||                                |||
  1142 ..SerValIleSerSer.........................SerThr 1148

  1928 TCCCTCTTCACCAACGTATCGAGTTCTTCCAAGTAAACTTCACCGTTACC 1879
       :::      |||::::::   |||:::|||:::   :::|||   :::||
  1149 ThrThrSerThrSerIlePheSerGluSerSerLysSerSerValIlePr 1165

  1878 GTACAACAACTCCACGGGGAACATAAGAGCTATGTCGAAAACCAGCGCAC 1829
       |   ::::::||||||::::::      :::   |||:::||||||:::
  1165 oThrSerSerSerThrSerGly...SerSerGluSerGluThrSerSer. 1180

  1828 CCTTCTTATCCATAGCCTCTGCTACCTGCTGCGGGGATATATCGTACACA 1779
                                  :::|||::::::|||   :::
  1181 ..........................AlaGlySerValSerSerSer 1187

  1778 TCCACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCG 1729
       |||   |||:::      |||:::   :::      :::   :::|||
  1188 SerPheIleSerSerGluSerSerLysSerProThrTyrSerSerSerSe 1204

  1728 ACTACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCACTGG 1679
         |||         |||                |||:::      |||   :
  1204 rLeuProLeuValThr..............SerAlaThrThrSerGlnG 1216

  1678 ACACGTACGAATCT...CCCCCAGCTGTGGAAAGAAAG............ 1644
       ::|||   :::|||   |||||||||:::      |||
  1216 luThrAlaSerSerLeuProProAlaThrThrThrLysThrSerGluGln 1232

  1643 ..............................................AG 1642

  1233 ThrThrLeuValThrValThrSerCysGluSerHisValCysThrGluSe 1249

  1641 GATCTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGC 1592
         ||||||      :::      |||:::   ::::::   :::|||:::
  1249 rIleSerProAlaIleValSerThrAlaThrValThrValSerGlyValT 1266

  1591 ATACATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCC 1542
         |||         ||||||
  1266 hrThrGluTyrThrThrTrp.........................Cys 1273

  1541 CCAATATCGAAGAAGCACACA 1521
       |||||||||      :::|||
  1274 ProIleSerThrThrGluThr 1280
```

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-478-507-7

seq_documentation_block:
; Sequence 7, Application US/08478507

```
; Patent No. 6120988
;  GENERAL INFORMATION:
;    APPLICANT:  Reyes, Gregory R
;    APPLICANT:  Yarbough, Patrice O
;    APPLICANT:  Bradley, Daniel W
;    APPLICANT:  Krawczynski, Krzysztof Z
;    APPLICANT:  Tam, Albert
;    APPLICANT:  Fry, Kirk E
;    TITLE OF INVENTION:  DNA Sequences of Enterically Transmitted
;    TITLE OF INVENTION:  No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,507
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/279,823
;      FILING DATE:  25-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,078
;      FILING DATE:  05-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/505,888
;      FILING DATE:  05-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/420,921
;      FILING DATE:  13-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/367,486
;      FILING DATE:  16-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/336,672
;      FILING DATE:  11-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/208,997
;      FILING DATE:  17-JUN-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0183.22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 324-0880
;      TELEFAX:  (650) 324-0960
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-478-507-7


alignment_scores:
            Quality:  169.50              Length:    890
              Ratio:  0.467                 Gaps:     41
 Percent Similarity:  40.787   Percent Identity: 20.112

alignment_block:
 US-09-301-906-3 x US-08-478-507-7   ..

 Align seg 1/1  to: US-08-478-507-7  from: 1  to: 1693
```

```
4561 GGTAAACAGCTTTGCATGTTTAGAACGTGT............TCCTGCGG 4598
     |||:::||||||   ::::::      |||           |||:::||
 425 GlyArgGlnLeuGluPheTyrAlaGlnCysArgArgTrpLeuSerAlaGl 441

4599 TGTGCAGCTGGAT............GTGTACAACGAAGCG.......... 4626
     |   :::||||||            |||::::::|||:::
 441 yPheHisLeuAspProArgValLeuValPheAspGluSerAlaProCysH 458

4627 ........ACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
             |||         :::      :::   |||
 458 isCysArgThrAlaIleArgLysAlaLeuSerLysPheCysCysPheMet 474

4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGG.. 4716
     :::      |||:::::::::::   |||   :::   :::||||||
 475 LysTrpLeuGlyGlnGluCysThrCysPheLeuGlnProAlaGluGlyAl 491

4717 ........................TATACCTATAATGGT.......... 4731
                              :::|||:::|||
 491 aValGlyAspGlnGlyHisAspAsnGluAlaTyrGluGlySerAspValA 508

4732 ..............................GGTAGCCATGTTTCATCA 4749
                                    ||||||:::|||
 508 spProAlaGluSerAlaIleSerAspIleSerGlySerTyrValValPro 524

4750 GGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAG 4799
     |||         ||||||:::   :::      |||:::      |||::
 525 Gly......ThrAlaLeuGlnProLeuTyrGlnAlaLeuAspLeuProAl 539

4800 CGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTAC 4849
     :         :::   :::::::::      ::::::   :::   |||
 539 a.........GluIleValAlaArgAlaGlyArgLeuThrAlaThrVal. 552

4850 CATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTG 4899
                        :::             |||:::   |||
 553 ...................LysValSerGlnValAspGlyArgIleAsp 562

4900 ACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGG 4949
          :::|||:::|||:::   :::|||   |||:::
 563 CysGluThrLeuLeuGlyAsnLysThrPheArgThrSerPhe...ValAs 578

4950 TGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGC 4999
      |||   |||:::      |||
 578 pGlyAlaValLeuGluThrAsn........................... 585

5000 CTTGCGGTTTTCAAAGGACGCACTTGCAT.........TCAGTAAACTCC 5040
          |||   :::|||   :::|||             |||   ::::::
 586 .....GlyProGluArgHisAsnLeuSerPheAspAlaSerGlnSerThr 600

5041 ATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC..... 5085
     :::      |||   :::||||||||||:::   |||:::
 601 MetAlaAlaGlyProPheSerLeuThrTyrAlaAlaSerAlaAlaGlyLe 617

5086 ..............GTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
                     :::   |||:::   :::         :::|||
 617 uGluValArgTyrValAlaAlaGlyLeuAspHisArgAlaValPheAlaP 634

5120 GCGGCGTGTCGGATGAGAAGTCACCAGGT..................... 5148
       |||||||||   :::::::::||||||
 634 roGlyValSerProArgSerAlaProGlyGluValThrAlaPheCysSer 650

5149 .........................................GTTCCAAA 5156
                                               :::   ||
 651 AlaLeuTyrArgPheAsnArgGluAlaGlnArgHisSerLeuIleGlyAs 667

5157 CCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGG 5206
     |         :::   :::|||            ::::::|||   |||
 667 nLeuTrpPheHisProGluGlyLeuIleGlyLeuPheAlaProPheSerP 684

5207 GGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCG 5256
       |||   :::   :::      |||::::::         |||:::|||
 684 roGlyHisValTrpGlu......SerAlaAsnProPheCysGlyGluSer 698

5257 ACATAC......TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
```

```
        |||              |||   |||   |||   |||   :::
 699 ThrLeuTyrThrArgThrTrpSerGluValAspAlaValSerSerProAl 715

5301 GTGG.................................CTCAGGGCTG 5314
                                           ||||||
 715 aArgProAspLeuGlyPheMetSerGluProSerIleProSerArgAlaA 732

5315 ATAATCCAGTGATGGCTCTT..............GAACCTGACTACACC 5349
       :::|||::::::|||                  ||||||   :::
 732 laThrProThrLeuAlaAlaProLeuProProProAlaProAspProSer 748

5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
     |||                        |||      |||:::::
 749 Pro.......................ProProSerAlaProAlaLeuAl 757

5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGT 5449
     :|||      :::      :::   |||               :::
 757 aGluProAlaSerGlyAlaThrAlaGlyAlaProAlaIleThrHisGlnT 774

5450 TGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAA 5499
               |||::::::      |||      :::|||
 774 hrAlaArgHisArgArgLeuLeuPheThrTyrProAspGlySerLysVal 790

5500 GTACCTAATCAAGTTTATGAATCACTA....................... 5526
          :::::::::::::||||||
 791 PheAlaGlySerLeuPheGluSerThrCysThrTrpLeuValAsnAlaSe 807

5527 ..............CCGGGC.....................TTTCACG 5539
                    ||||||                    ::::::
 807 rAsnValAspHisArgProGlyGlyGlyLeuCysHisAlaPheTyrGlnA 824

5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGC... 5586
       |||   ::::::   |||                  :::||||||
 824 rgTyrProAlaSerPheAspAlaAlaSerPheValMetArgAspGlyAla 840

5587 ............TTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGA 5624
                |||       |||   :::   :::::::::      ::
 841 AlaAlaTyrThrLeuThrProArgProIleIleHisAlaValAlaProAs 857

5625 AAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGG 5674
     :      :::      :::   |||   ::::::|||
 857 pTyrArgLeuGluHisAsnProLysArgLeuGluAla.............. 869

5675 GCGACAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATT 5724
                                 ::::::::::::::
 870 ..............................AlaTyrArgGluThrCys 875

5725 AATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATG........... 5763
     :::   :::|||:::                |||
 876 SerArgLeuGlyThrAlaAlaTyrProLeuLeuGlyThrGlyIleTyrGl 892

5764 .......GTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATG 5806
             :::|||   ||||||   :::::::|||            |||
 892 nValProIleGlyProSerPheAspAlaTrpGlu.........ArgAsnH 906

5807 CTCCCCCAGGTGGC................................. 5820
       ||||||
 906 isArgProGlyAspGluLeuTyrLeuProGluLeuAlaAlaArgTrpPhe 922

5821 ..................GGTAAGACGACGACGTTAGTGGACGAATTCGT 5852
                         |||   |||::::::::::::   ::
 923 GluAlaAsnArgProThrArgProThrLeuThrIleThrGluAspValAl 939

5853 TAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTG 5902
     ::::            ||||||:::::::::   ::::::   |||:::
 939 aArg............ThrAlaAsnLeuAlaIleGluLeuAspSerAlaT 952

5903 AGGACATAAATATGGCGGTGAAG........................... 5925
       |||::::::   |||
 952 hrAspValGlyArgAlaCysAlaGlyCysArgValThrProGlyValVal 968

5926 ......AAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGT 5969
           :::         ::::::   |||   :::::::::   :::::
```

```
 969 GlnTyrGlnPheThrAlaGlyValProGlySerGlyLysSerArgSerIl 985

5970 TAACTCCAGGGTGGTAAACTTTATCGTC..................... 5997
     :::::::      |||:::   :::|||
 985 eThrGlnAlaAspValAspValValValValProThrArgGluLeuArgA 1002

5998 .............AGGGGAATGTAT..................... 6009
                  ||||||:::
1002 snAlaTrpArgArgArgGlyPheAlaAlaPheThrProHisThrAlaAla 1018

6010 .............AAAAGGGTTTTGGTGGATGAGGTGCACATGATGCA 6044
                  :::||||||::::::||||||:::      :::
1019 ArgValThrGlnGlyArgArgValValIleAspGluAlaProSerLeuPr 1035

6045 TCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCC 6094
           ||||||   |||   :::           |||:::
1035 oProHisLeuLeuLeuLeuHisMet......GlnArgAlaAlaThrValH 1050

6095 TCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTG 6144
       ::::::||||||   ||||||||||||   |||:::   |||
1050 isLeuLeuGlyAspProAsnGlnIleProAlaIleAspPheGluHis... 1065

6145 TTTAGGATGGATTGTGCTGTTTTTGTTCCAAAG...AAGGAAAGCGTTGT 6191
                  |||   :::||||||      :::   ::::::
1066 .............AlaGlyLeuValProAlaIleArgProAspLeuGl 1077

6192 ATACACTTCT.........AAATCGTACAGGTGTCCGTTAGATGTTTGCT 6232
         ||||||           ::::::|||   |||   |||||||||
1077 yProThrSerTrpTrpHisValThrHisArgTrpProAlaAspValCysG 1094

6233 ACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAA 6282
       |||                :::||||||        :::||||||
1094 luLeu.............IleArgGly.........AlaTyrProMet 1102

6283 AAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTG......TC 6326
           ::::::::::::         |||:::||||||||||      ::
1103 IleGlnThrThrSerArg.........ValLeuArgSerLeuPheTrpGl 1116

6327 CAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACG 6376
     ::::      :::|||
1116 yGluProAlaValGlyGlnLys........................... 1123

6377 TGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTG 6426
           |||   :::||||||      |||                |||
1124 .....LeuValPheThrGlnAlaAlaLysProAlaAsnProGlySer... 1137

6427 AAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGG 6476
                                 :::||||||||||||||||||||
1138 ..........................ValThrValHisGluAlaGlnGl 1145

6477 AAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACT 6526
     |   |||:::::::::::::::::::::   |||      |||
1145 yAlaThrTyrThrGluThrThrIleIleAlaThrAlaAspAlaArgGlyL 1162

6527 CCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
       :::   :::::::::   |||   :::|||:::|||:::||||||||||
1162 euIleGlnSerSerArgAlaHisAlaIleValAlaLeuThrArgHisThr 1178

6577 CGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGG 6626
     ::::::   |||                :::   |||   :::   |||||
1179 GluLysCysValIleIleAspAlaProGlyLeuLeuArgGlu...ValGl 1194

6627 CACATATATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTC 6676
     |                              :::||||||||:::::::
1194 y..............................IleSerAspAlaIleValA 1201

6677 ACACGTTCGCCCCGGCTGGT 6696
     :::::|||      ||||||
1201 snAsnPhePheLeuAlaGly 1207

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-616-844-40

seq_documentation_block:

; Sequence 40, Application US/08616844
; Patent No. 5849578
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A.
;    TITLE OF INVENTION:  COMPOSITION AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/616,844
;      FILING DATE:  15-MAR-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,654
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,573
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-053
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-616-844-40


alignment_scores:
            Quality:  161.50              Length:   1318
              Ratio:  0.273                 Gaps:     67
 Percent Similarity:  44.917   Percent Identity: 20.865

alignment_block:
 US-09-301-906-3/rev x US-08-616-844-40   ..

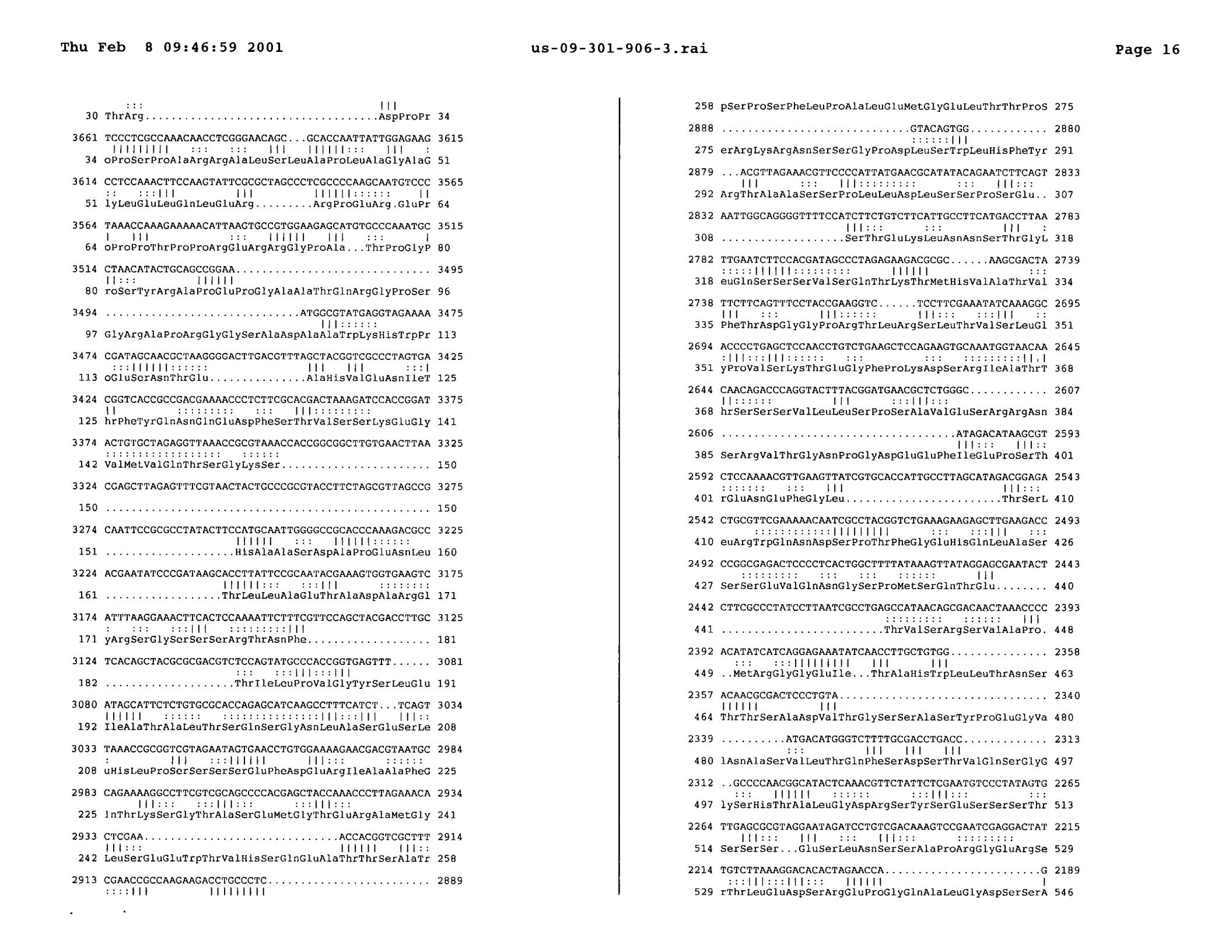
 Align seg 1/1  to: US-08-616-844-40  from: 1  to: 1481

   3852 ATGCTGAGTCCC.........AGATGGACCAGCACCTTCGTCATAGTCAA 3812
        |||   ||||||         ||||||         ::::::::::::
      1 MetAlaSerProArgAlaSerArgTrpProProProLeuLeuLeuLeLe 17

   3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCCTCTACC 3762
         :::|||                   ::::::   :::||||||
     17 uLeuPro.................LeuLeuLeuLeuProPro....... 25

   3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTAT 3712
                                                 ||||||
     26 .....................................AlaAlaProGly 29

   3711 GATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662
```




2188 CTACAACAGGTATGAACACCAGCGATTTGCCCACGAACGACGGGACCAAC 2139
546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561
2138 GACCTATACGTAACCTCAGT.....TCCAGGTGACTTCTGAGCCCTAC 2096
562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578
2095 ACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTCTAGCTTA 2046
578 rLeuThrAsnThrSerMetSerThrThr..... 587
2045 AACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAA.....AAA 2002
588 SerGlyGluAlaGlySer 593
2001 CCCGCTTACGTT.....AGAGAAGGAATGTTCATACATCTCAC 1964
594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609
1963 CACACTGACCAACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914
610 .AsnValThrAspAspMetGlyLeuVal.....SerA 620
1913 GT.....ATCGAGTTCTTC 1900
620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636
1899 CAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACAT..... 1855
637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653
1854 AAGAGCTATGTCGAAAACCAGCGCACCCTTCTTA 1821
653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669
1820 TCCATAGCCTCTGCTACCTGCTGCGGGGATATA.....TCGTACACATC 1777
670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrSe 686
1776 CACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCGAC 1727
686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr..... 698
1726 TACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCA...CTG 1680
699 SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711
1679 GACACGTACGAA.....TCTCCCCCAGCTGTGGAAAGAAAGAGGAT 1639
712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu..... 724
1638 CTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGCATA 1589
725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr..... 737
1588 CATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCCCCA 1539
738 ProThr 739
1538 ATATCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAA 1489
740 IleAsnMetProAsnThr..... 745
1488 ACAGCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGACCGAGTCGG 1439
746 SerValValLeuAspThrAspAlaG 754
1438 AGAACTGTATCGAAAGCTCTGGAAAGAGTTCCCTCAGCTTCT..... 1397
754 luPheValSerAspSerSerSerSerSerSer.SerSerSerSerSerSe 770
1396 CCCTTTCCCGCGGCTTCAGAAACACATCGCA 1366
770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis..Hi 786
1365 CCTATGTGAGGGTTTTTCGCGAATCATTGCGTCCCTTTGAGAAAACATAT 1316
786 sLeu.....PheSerSerIleLeuProSerThrArgAlaSerVal. 799
1315 CGCACAATTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAAT 1266
800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815
1265 ACCAGTCTATTGGTGATACTCGAAACGACCGATGAATAGACCTCCTCTTG 1216
816 ProSerProLeuProValSerLeuThrThrSer.....ThrSerAlaPr 830
1215 ATTAAGGGCATTGGCGACGGAGGCCAAGATCTCCCCTCCCAACCGTGTCC 1166
830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847
1165 TATCGCCAGCGACGATAGCGGTTTTTGGGACATTGGGGAGATAGACGTCT 1116
847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863
1115 CCGCCCTTCCACCCAAAGGTACGCTCTTCATTGAGTGATGGCATGTGTAG 1066
864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880
1065 TATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAG...CCAA 1019
880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897
1018 GGTGCATCAAGAAACCTATCGAAAAG..... 993
897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913
992 GAAGATGAGAAGCAACGACGTCTGTAAAT 964
914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928
963 TTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTGGTGGTGC 914
929 ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944
913 ATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGGCAAGGCA 864
944 hrThr.....IleLeuLysThrSerGlnProLeuMetThr 955
863 ACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATC..... 822
956 Thr.....ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970
821 .GTAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGA 773
970 oIleAlaValGlnThrThrAlaGlyLysGln.....LeuSerLeuT 984
772 CGTAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAG.....ATTTTA 729
984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000
728 GCCCAAACGAACGTGTTCTTTACATCAATAGGAACATTAAGGGCCCTGTT 679
1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017
678 AAACAATTTCCCT.....TTCCGCAAGATGACAGCGTCGCTCTCA.... 639
1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034
638 TCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAG 600
1034 hrAlaGluTyrSerPro.....AlaSerArg.SerLeuGlyTh 1046
599 GCGGCCTTCTCCTCCAACGACTGTTGTGGGACT.....GATTGCGCTG 557

```
          |||||||||   ||||||||||||:::         |||   |||
    1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063

     556 GCTCCAGTGCTTCCTGTTTTTTAGGAAGTGGTACGTGGACGCCAACAGCT 507
         :::|||||:::      :::   |||:::|||   :::||||||:::
    1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079

     506 GTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGCTTCTGC 462
         ::::::   :::|||:::   ::::::|||   ::::::   |||
    1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-599-654-40

seq_documentation_block:
; Sequence 40, Application US/08599654
; Patent No. 5882925
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,654
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,573
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-599-654-40


alignment_scores:
           Quality:  161.50            Length:   1318
             Ratio:  0.273               Gaps:     67
 Percent Similarity:  44.917   Percent Identity: 20.865

alignment_block:
 US-09-301-906-3/rev x US-08-599-654-40   ..

 Align seg 1/1  to: US-08-599-654-40  from: 1  to: 1481

    3852 ATGCTGAGTCCC.........AGATGGACCAGCACCTTCGTCATAGTCAA 3812
         |||   ||||||         ||||||         ::::::::::::

    1 MetAlaSerProArgAlaSerArgTrpProProProLeuLeuLeuLeuLe 17

 3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCCTCTACC 3762
       :::|||                  ::::::   :::||||||
   17 uLeuPro.................LeuLeuLeuLeuProPro....... 25

 3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTAT 3712
                                                 ||||||
   26 .....................................AlaAlaProGly 29

 3711 GATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662
         :::                                     |||
   30 ThrArg.....................................AspProPr 34

 3661 TCCCTCGCCAAACAACCTCGGGAACAGC...GCACCAATTATTGGAGAAG 3615
       |||||||||   :::   :::   |||   ||||||:::   |||   :
   34 oProSerProAlaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51

 3614 CCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCAAGCAATGTCCC 3565
      ::   :::|||         |||         ||||||::::::    ||
   51 lyLeuGluLeuGlnLeuGluArg.........ArgProGluArg.GluPr 64

 3564 TAAACCAAAGAAAAACATTAAGTGCCGTGGAAGAGCATGTGCCCAAATGC 3515
      |   |||            :::   ||||||   |||   :::      |
   64 oProProThrProProArgGluArgArgGlyProAla...ThrProGlyP 80

 3514 CTAACATACTGCAGCCGGAA.............................. 3495
      ||:::         ||||||
   80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96

 3494 ............................ATGGCGTATGAGGTAGAAAA 3475
                                  |||::::::
   97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113

 3474 CGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACGGTCGCCCTAGTGA 3425
       :::||||||::::::              |||   |||      :::|
  113 oGluSerAsnThrGlu..............AlaHisValGluAsnIleT 125

 3424 CGGTCACCGCCGACGAAAACCCTCTTCGCACGACTAAAGATCCACCGGAT 3375
      ||         :::::::::   :::   |||::::::::::
  125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141

 3374 ACTGTGCTAGAGGTTAAACCGCGTAAACCACCGGCGGCTTGTGAACTTAA 3325
      ::::::::::::::::::   ::::::
  142 ValMetValGlnThrSerGlyLysSer...................... 150

 3324 CGAGCTTAGAGTTTCGTAACTACTGCCCGCGTACCTTCTAGCGTTAGCCG 3275

  150 ..................................................... 150

 3274 CAATTCCGCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAAGACGCC 3225
                        ||||||   :::   ||||||::::::
  151 ...................HisAlaAlaSerAspAlaProGluAsnLeu 160

 3224 ACGAATATCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGTGAAGTC 3175
                      ||||||:::   :::|||      ::::::::
  161 .................ThrLeuLeuAlaGluThrAlaAspAlaArgGl 171

 3174 ATTTAAGGAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGACCTTGC 3125
      :   :::   :::|||   :::::::::|||
  171 yArgSerGlySerSerSerArgThrAsnPhe................... 181

 3124 TCACAGCTACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTT...... 3081
                        :::   :::|||:::|||
  182 ...................ThrIleLeuProValGlyTyrSerLeuGlu 191

 3080 ATAGCATTCTCTGTGCGCACCAGAGCATCAAGCCTTTCATCT...TCAGT 3034
      ||||||   ::::::   :::::::::::::|||:::|||   |||::
  192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208

 3033 TAAACCGCGGTCGTAGAATAGTGAACCTGTGGAAAAGAACGACGTAATGC 2984
      :         |||   :::||||||      |||:::      ::::::
  208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225
```

```
2983 CAGAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAACCCTTAGAAACA 2934
225 lnThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241

2933 CTCGAA..........................ACCACGGTCGCTTT 2914
242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258

2913 CGAACCGCCAAGAAGACCTGCCCTC..................... 2889
258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275

2888 ..........................GTACAGTGG............ 2880
275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291

2879 ...ACGTTAGAAACGTTCCCCATTATGAACGCATATACAGAATCTTCAGT 2833
292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu.. 307

2832 AATTGGCAGGGGTTTTCCATCTTCTGTCTTCATTGCCTTCATGACCTTAA 2783
308 ..................SerThrGluLysLeuAsnAsnSerThrGlyL 318

2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCGC......AAGCGACTA 2739
318 euGlnSerSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334

2738 TTCTTCAGTTTCCTACCGAAGGTC......TCCTTCGAAATATCAAAGGC 2695
335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGl 351

2694 ACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAAGTGCAAATGGTAACAA 2645
351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368

2644 CAACAGACCCAGGTACTTTACGGATGAACGCTCTGGGC............ 2607
368 hrSerSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384

2606 ..................................ATAGACATAAGCGT 2593
385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401

2592 CTCCAAAACGTTGAAGTTATCGTGCACCATTGCCTTAGCATAGACGGAGA 2543
401 rGluAsnGluPheGlyLeu.....................ThrSerL 410

2542 CTGCGTTCGAAAAACAATCGCCTACGGTCTGAAAGAAGAGCTTGAAGACC 2493
410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426

2492 CCGGCGAGACTCCCCTCACTGGCTTTTATAAAGTTATAGGAGCGAATACT 2443
427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu........ 440

2442 CTTCGCCCTATCCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393
441 .......................ThrValSerArgSerValAlaPro. 448

2392 ACATATCATCAGGAGAAATATCAACCTTGCTGTGG............... 2358
449 ..MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463

2357 ACAACGCGACTCCCTGTA............................. 2340
464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480

2339 ..........ATGACATGGGTCTTTTGCGACCTGACC............. 2313
480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497

2312 ..GCCCCAACGGCATACTCAAACGTTCTATTCTCGAATGTCCCTATAGTG 2265
497 lySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThr 513

2264 TTGAGCGCGTAGGAATAGATCCTGTCGACAAAGTCCGAATCGAGGACTAT 2215
514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529

2214 TGTCTTAAAGGACACACTAGAACCA.......................G 2189
529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerA 546

2188 CTACAACAGGTATGAACACCAGCGATTTGCCCACGAACGACGGGACCAAC 2139
546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561

2138 GACCTATACGTAACCTCAGT.......TCCAGGTGACTTCTGAGCCCTAC 2096
562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578

2095 ACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTCTAGCTTA 2046
578 rLeuThrAsnThrSerMetSerThrThr..................... 587

2045 AACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAA......AAA 2002
588 ..............................SerGlyGluAlaGlySer 593

2001 CCCGCTTACGTT............AGAGAAGGAATGTTCATACATCTCAC 1964
594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609

1963 CACACTGACCAACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914
610 .AsnValThrAspAspMetGlyLeuVal.................SerA 620

1913 GT..................................ATCGAGTTCTTC 1900
620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636

1899 CAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACAT..... 1855
637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653

1854 ..............AAGAGCTATGTCGAAAACCAGCGCACCCTTCTTA 1821
653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669

1820 TCCATAGCCTCTGCTACCTGCTGCGGGGATATA......TCGTACACATC 1777
670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrSe 686

1776 CACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCGAC 1727
686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr............. 698

1726 TACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCA...CTG 1680
699 ..........SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711

1679 GACACGTACGAA.........TCTCCCCCAGCTGTGGAAAGAAAGAGGAT 1639
712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu........... 724

1638 CTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGCATA 1589
725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr.......... 737

1588 CATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCCCCA 1539
738 ..........................................ProThr 739

1538 ATATCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAA 1489
```

```
     |||:::      :::|||
  740 IleAsnMetProAsnThr.............................. 745

 1488 ACAGCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGACCGAGTCGG 1439
                              :::   :::|||   |||:::::::|
  746 .......................SerValValLeuAspThrAspAlaG 754

 1438 AGAACTGTATCGAAAGCTCTGGAAAGAGTTCCCTCAGCTTCT........ 1397
     ||         :::||||||:::::||||||  |||:::|||
  754 luPheValSerAspSerSerSerSerSerSer.SerSerSerSerSerSe 770

 1396 ..................CCCTTTCCCGCGGCTTCAGAAACACATCGCA 1366
                        |||:::|||:::::::|||:::::::|||  ||
  770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis..Hi 786

 1365 CCTATGTGAGGGTTTTTCGCGAATCATTGCGTCCCTTTGAGAAAACATAT 1316
     ||||         ||||||   |||:::   |||         ::::::
  786 sLeu.........PheSerSerIleLeuProSerThrArgAlaSerVal. 799

 1315 CGCACAATTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAAT 1266
        |||   ::   |||      ::::::::::::   ||||||   :::
  800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815

 1265 ACCAGTCTATTGGTGATACTCGAAACGACCGATGAATAGACCTCCTCTTG 1216
        |||  |||   :::      ||||||:::      ||||||:::
  816 ProSerProLeuProValSerLeuThrThrSer......ThrSerAlaPr 830

 1215 ATTAAGGGCATTGGCGACGGAGGCCAAGATCTCCCCTCCCAACCGTGTCC 1166
      |||   :::      |||          |||      :::   ||||
  830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847

 1165 TATCGCCAGCGACGATAGCGGTTTTTGGGACATTGGGGAGATAGACGTCT 1116
     ||      |||      :::         :::::         |||
  847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863

 1115 CCGCCCTTCCACCCAAAGGTACGCTCTTCATTGAGTGATGGCATGTGTAG 1066
           |||         ::::::||||||   :::      :::   ||
  864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880

 1065 TATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAG...CCAA 1019
     |:::   :::|||   :::      ::::::::::::   |||   |||
  880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897

 1018 GGTGCATCAAGAAACCTATCGAAAAG...................... 993
           ::::::   |||   |||:::
  897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913

  992 .....................GAAGATGAGAAGCAACGACGTCTGTAAAT 964
                           |||::::::   ::::::   |||
  914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928

  963 TTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTGGTGGTGC 914
         |||   :::   :::   |||   |||      |||      :::
  929 ....ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944

  913 ATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGGCAAGGCA 864
        |||               ||||||   :::|||:::         :::
  944 hrThr..............IleLeuLysThrSerGlnProLeuMetThr 955

  863 ACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATC........ 822
     |||      ||||||::::::   :::      |||   :::
  956 Thr......ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970

  821 .GTAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGA 773
       :::|||:::      |||   ||||||:::         ||||||||||
  970 oIleAlaValGlnThrThrAlaGlyLysGln.........LeuSerLeuT 984

  772 CGTAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAG......ATTTTA 729
     ||            |||:::   :::   |||:::::::      |||
  984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000

  728 GCCCAAACGAACGTGTTCTTTACATCAATAGGAACATTAAGGGCCCTGTT 679
     ::::::   |||      :::::::::      |||            ||
```

```
     1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017

      678 AAACAATTTCCCT......TTCCGCAAGATGACAGCGTCGCTCTCA.... 639
          |:::::::   |||         ::::::||||||::::::|||:::
     1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034

      638 ...........TCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAG 600
                     ||||||              :::||| :::   :::
     1034 hrAlaGluTyrSerPro............AlaSerArg.SerLeuGlyTh 1046

      599 GCGGCCTTCTCCTCCAACGACTGTTGTGGGACT.......GATTGCGCTG 557
             |||||||||   ||||||||||||:::         |||   |||
     1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063

      556 GCTCCAGTGCTTCCTGTTTTTTAGGAAGTGGTACGTGGACGCCAACAGCT 507
          :::||||||:::      :::   |||:::|||   :::||||||:::
     1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079

      506 GTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGCTTCTGC 462
          ::::::   :::|||:::   :::::::|||   ::::::   |||
     1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094
```

```
seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-868A-40

seq_documentation_block:
; Sequence 40, Application US/08944868A
; Patent No. 6018025
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,868A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/599,654
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-041
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-944-868A-40


alignment_scores:
```

```
          Quality:  161.50              Length:   1318
            Ratio:  0.273                Gaps:     67
Percent Similarity:  44.917   Percent Identity: 20.865

alignment_block:
US-09-301-906-3/rev x US-08-944-868A-40   ..

Align seg 1/1  to: US-08-944-868A-40  from: 1  to: 1481

   3852 ATGCTGAGTCCC.........AGATGGACCAGCACCTTCGTCATAGTCAA 3812
          |||   ||||||         ||||||         ::::::::::::
      1 MetAlaSerProArgAlaSerArgTrpProProProLeuLeuLeuLeLe 17

   3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCCTCTACC 3762
         :::|||                 ::::::   :::||||||
     17 uLeuPro.................LeuLeuLeuLeuProPro....... 25

   3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTAT 3712
                                                  ||||||
     26 .....................................AlaAlaProGly 29

   3711 GATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662
           :::                                      |||
     30 ThrArg.................................AspProPr 34

   3661 TCCCTCGCCAAACAACCTCGGGAACAGC...GCACCAATTATTGGAGAAG 3615
         |||||||||   :::   :::   |||   ||||||:::   |||   :
     34 oProSerProAlaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51

   3614 CCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCAAGCAATGTCCC 3565
         ::   :::|||         |||         ||||||::::::    ||
     51 lyLeuGluLeuGlnLeuGluArg.........ArgProGluArg.GluPr 64

   3564 TAAACCAAAGAAAAACATTAAGTGCCGTGGAAGAGCATGTGCCCAAATGC 3515
         |   |||            :::   ||||||   |||   :::      |
     64 oProProThrProProArgGluArgArgGlyProAla...ThrProGlyP 80

   3514 CTAACATACTGCAGCCGGAA.............................. 3495
         ||:::          ||||||
     80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96

   3494 ...........................ATGGCGTATGAGGTAGAAAA 3475
                                    |||::::::
     97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113

   3474 CGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACGGTCGCCCTAGTGA 3425
         :::||||||:::::               |||   |||      :::|
    113 oGluSerAsnThrGlu.............AlaHisValGluAsnIleT 125

   3424 CGGTCACCGCCGACGAAAACCCTCTTCGCACGACTAAAGATCCACCGGAT 3375
         ||       :::::::::   :::   |||:::::::::
    125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141

   3374 ACTGTGCTAGAGGTTAAACCGCGTAAACCACCGGCGGCTTGTGAACTTAA 3325
         ::::::::::::::::   ::::::
    142 ValMetValGlnThrSerGlyLysSer..................... 150

   3324 CGAGCTTAGAGTTTCGTAACTACTGCCCGCGTACCTTCTAGCGTTAGCCG 3275

    150 ..................................................... 150

   3274 CAATTCCGCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAAGACGCC 3225
                          ||||||   :::   ||||||::::::
    151 ...................HisAlaAlaSerAspAlaProGluAsnLeu 160

   3224 ACGAATATCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGTGAAGTC 3175
                           ||||||:::   :::|||      ::::::::
    161 .................ThrLeuLeuAlaGluThrAlaAspAlaArgGl 171

   3174 ATTTAAGGAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGACCTTGC 3125
         :   :::   :::|||   :::::::::|||
    171 yArgSerGlySerSerSerArgThrAsnPhe..................... 181

   3124 TCACAGCTACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTT..... 3081
                              :::   :::|||:::|||
    182 ....................ThrIleLeuProValGlyTyrSerLeuGlu 191

   3080 ATAGCATTCTCTGTGCGCACCAGAGCATCAAGCCTTTCATCT...TCAGT 3034
         ||||||   ::::::   ::::::::::::::::|||:::|||   |||::
    192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208

   3033 TAAACCGCGGTCGTAGAATAGTGAACCTGTGGAAAAGAACGACGTAATGC 2984
         :         |||   :::||||||      |||:::      ::::::
    208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225

   2983 CAGAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAACCCTTAGAAACA 2934
              |||:::   :::|||:::      :::|||:::
    225 lnThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241

   2933 CTCGAA..............................ACCACGGTCGCTTT 2914
         |||:::                            ||||||   |||::
    242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258

   2913 CGAACCGCCAAGAAGACCTGCCCTC.......................... 2889
         ::::|||         |||||||||
    258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275

   2888 ............................GTACAGTGG............ 2880
                                    ::::::|||
    275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291

   2879 ...ACGTTAGAAACGTTCCCCATTATGAACGCATATACAGAATCTTCAGT 2833
           |||      :::   |||::::::::::      :::   |||:::
    292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu.. 307

   2832 AATTGGCAGGGGTTTTCCATCTTCTGTCTTCATTGCCTTCATGACCTTAA 2783
                            |||:::      :::          |||   :
    308 ..................SerThrGluLysLeuAsnAsnSerThrGlyL 318

   2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCGC......AAGCGACTA 2739
         :::::||||||:::::::::      ||||||          :::
    318 euGlnSerSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334

   2738 TTCTTCAGTTTCCTACCGAAGGTC......TCCTTCGAAATATCAAAGGC 2695
         |||   :::      |||::::::      |||:::   :::|||   ::
    335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGl 351

   2694 ACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAAGTGCAAATGGTAACAA 2645
         :|||:::|||::::::   :::          :::   :::::::::||||
    351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368

   2644 CAACAGACCCAGGTACTTTACGGATGAACGCTCTGGGC............. 2607
         ||::::::          |||      :::|||:::
    368 hrSerSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384

   2606 ................................ATAGACATAAGCGT 2593
                                        |||:::   |||::
    385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401

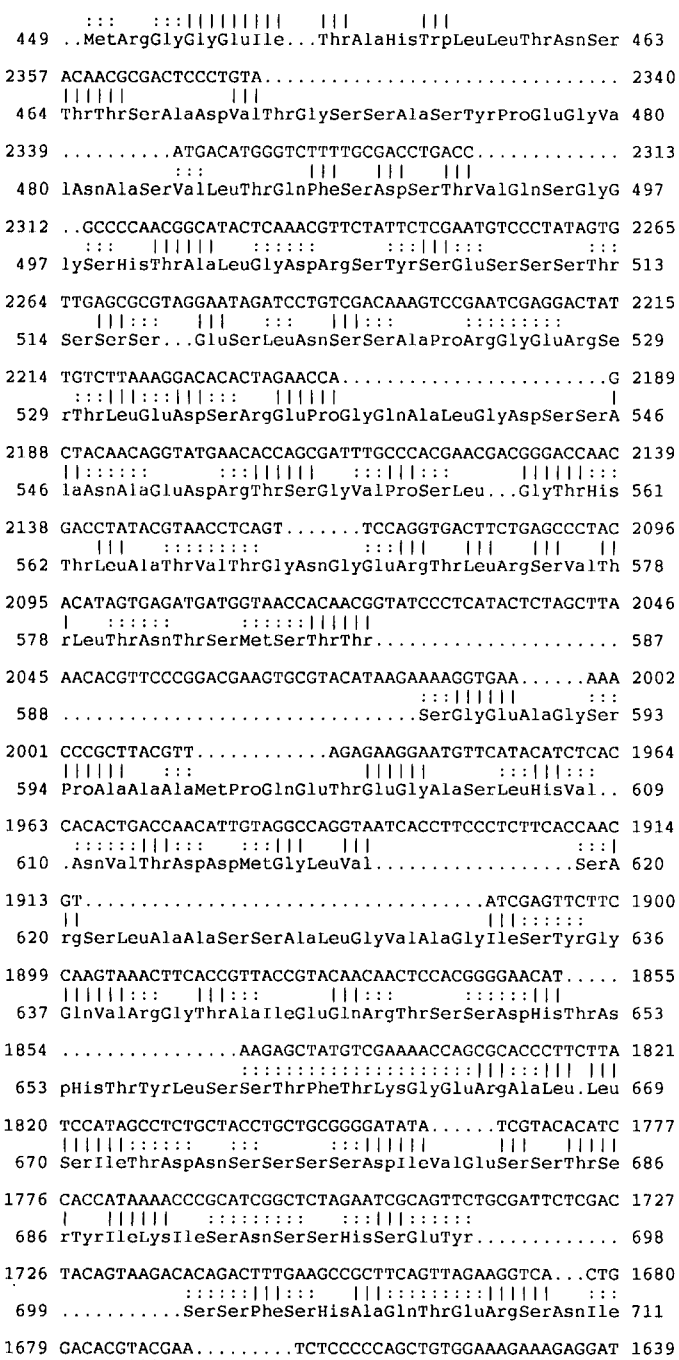
   2592 CTCCAAAACGTTGAAGTTATCGTGCACCATTGCCTTAGCATAGACGGAGA 2543
         :::::::   :::   |||                        |||:::
    401 rGluAsnGluPheGlyLeu.......................ThrSerL 410

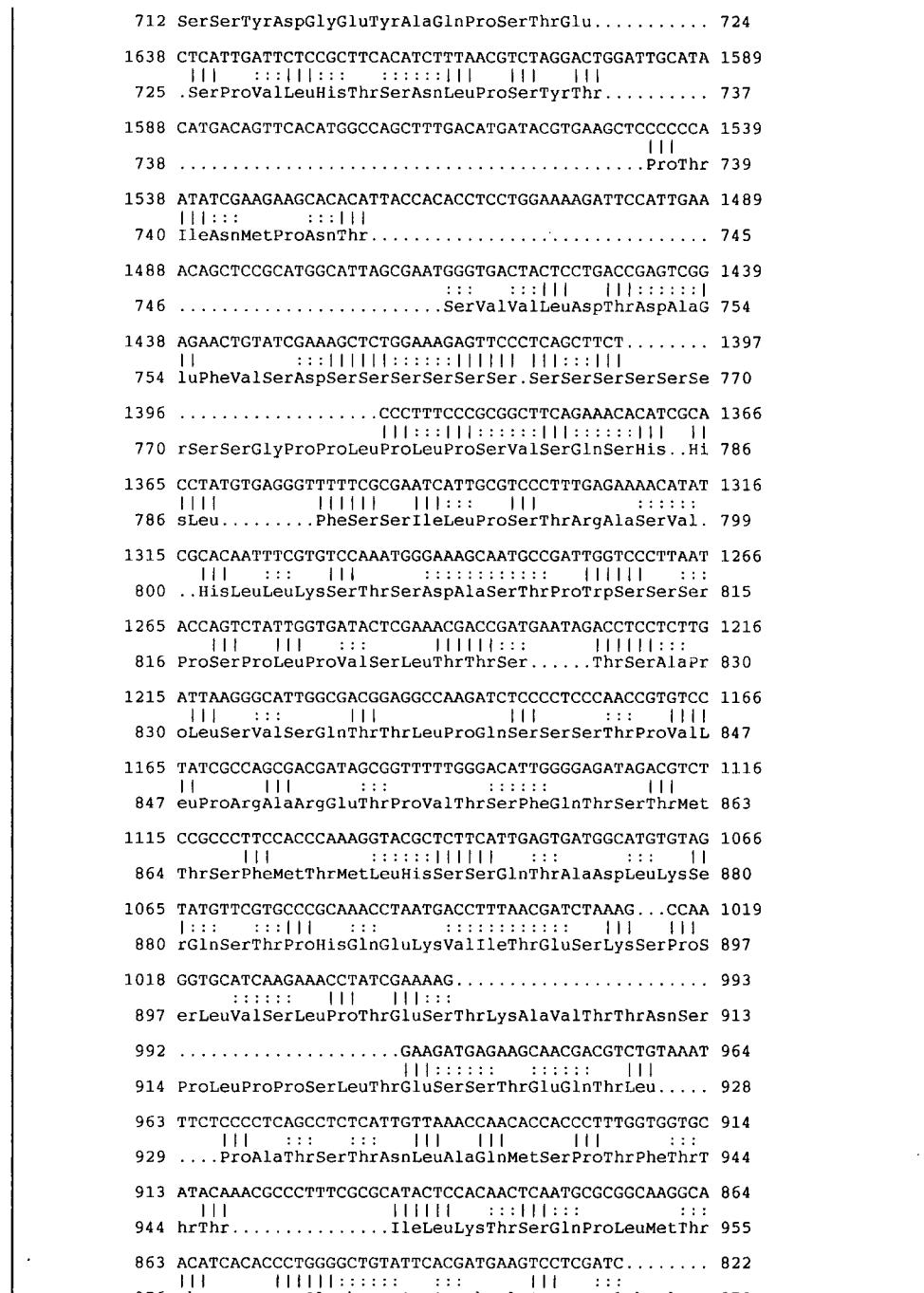
   2542 CTGCGTTCGAAAAACAATCGCCTACGGTCTGAAAGAAGAGCTTGAAGACC 2493
              ::::::::::::|||||||||      :::   :::|||   :::
    410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426

   2492 CCGGCGAGACTCCCCTCACTGGCTTTTATAAAGTTATAGGAGCGAATACT 2443
           :::::::::   :::   :::   ::::::      |||
    427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu........ 440

   2442 CTTCGCCCTATCCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393
                                  :::::::::   ::::::   |||
    441 ........................ThrValSerArgSerValAlaPro. 448

   2392 ACATATCATCAGGAGAAATATCAACCTTGCTGTGG............... 2358
```


449 ..MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463
2357 ACAACGCGACTCCCTGTA..... 2340
464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480
2339 ATGACATGGGTCTTTTGCGACCTGACC..... 2313
480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497
2312 ..GCCCCAACGGCATACTCAAACGTTCTATTCTCGAATGTCCCTATAGTG 2265
497 lySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThr 513
2264 TTGAGCGCGTAGGAATAGATCCTGTCGACAAAGTCCGAATCGAGGACTAT 2215
514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529
2214 TGTCTTAAAGGACACACTAGAACCA.....G 2189
529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerA 546
2188 CTACAACAGGTATGAACACCAGCGATTTGCCCACGAACGACGGGACCAAC 2139
546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561
2138 GACCTATACGTAACCTCAGT.....TCCAGGTGACTTCTGAGCCCTAC 2096
562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578
2095 ACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTCTAGCTTA 2046
578 rLeuThrAsnThrSerMetSerThrThr..... 587
2045 AACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAA.....AAA 2002
588 SerGlyGluAlaGlySer 593
2001 CCCGCTTACGTT.....AGAGAAGGAATGTTCATACATCTCAC 1964
594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609
1963 CACACTGACCAACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914
610 .AsnValThrAspAspMetGlyLeuVal.....SerA 620
1913 GT.....ATCGAGTTCTTC 1900
620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636
1899 CAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACAT..... 1855
637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653
1854 AAGAGCTATGTCGAAAACCAGCGCACCCTTCTTA 1821
653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669
1820 TCCATAGCCTCTGCTACCTGCTGCGGGGATATA.....TCGTACACATC 1777
670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrSe 686
1776 CACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCGAC 1727
686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr..... 698
1726 TACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCA...CTG 1680
699 SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711
1679 GACACGTACGAA.....TCTCCCCCAGCTGTGGAAAGAAAGAGGAT 1639


712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu..... 724
1638 CTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGCATA 1589
725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr..... 737
1588 CATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCCCCA 1539
738 ProThr 739
1538 ATATCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAA 1489
740 IleAsnMetProAsnThr..... 745
1488 ACAGCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGACCGAGTCGG 1439
746 SerValValLeuAspThrAspAlaG 754
1438 AGAACTGTATCGAAAGCTCTGGAAAGAGTTCCCTCAGCTTCT..... 1397
754 luPheValSerAspSerSerSerSerSerSer.SerSerSerSerSerSe 770
1396 CCCTTTCCCGCGGCTTCAGAAACACATCGCA 1366
770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis..Hi 786
1365 CCTATGTGAGGGTTTTTCGCGAATCATTGCGTCCCTTTGAGAAAACATAT 1316
786 sLeu.....PheSerSerIleLeuProSerThrArgAlaSerVal. 799
1315 CGCACAATTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAAT 1266
800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815
1265 ACCAGTCTATTGGTGATACTCGAAACGACCGATGAATAGACCTCCTCTTG 1216
816 ProSerProLeuProValSerLeuThrThrSer.....ThrSerAlaPr 830
1215 ATTAAGGGCATTGGCGACGGAGGCCAAGATCTCCCCTCCCAACCGTGTCC 1166
830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847
1165 TATCGCCAGCGACGATAGCGGTTTTTGGGACATTGGGGAGATAGACGTCT 1116
847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863
1115 CCGCCCTTCCACCCAAAGGTACGCTCTTCATTGAGTGATGGCATGTGTAG 1066
864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880
1065 TATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAG...CCAA 1019
880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897
1018 GGTGCATCAAGAAACCTATCGAAAAG..... 993
897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913
992 GAAGATGAGAAGCAACGACGTCTGTAAAT 964
914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928
963 TTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTGGTGGTGC 914
929 ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944
913 ATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGGCAAGGCA 864
944 hrThr.....IleLeuLysThrSerGlnProLeuMetThr 955
863 ACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATC..... 822
956 Thr.....ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970

```
     821 .GTAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGA 773
     970 oIleAlaValGlnThrThrAlaGlyLysGln.....LeuSerLeuT 984

     772 CGTAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAG......ATTTTA 729
     984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000

     728 GCCCAAACGAACGTGTTCTTTACATCAATAGGAACATTAAGGGCCCTGTT 679
    1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017

     678 AAACAATTTCCCT......TTCCGCAAGATGACAGCGTCGCTCTCA.... 639
    1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034

     638 ...........TCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAG 600
    1034 hrAlaGluTyrSerPro...........AlaSerArg.SerLeuGlyTh 1046

     599 GCGGCCTTCTCCTCCAACGACTGTTGTGGGACT.......GATTGCGCTG 557
    1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063

     556 GCTCCAGTGCTTCCTGTTTTTTAGGAAGTGGTACGTGGACGCCAACAGCT 507
    1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079

     506 GTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGCTTCTGC 462
    1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-423A-40

seq_documentation_block:
; Sequence 40, Application US/08944423A
; Patent No. 6020463
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  WINDOWS 95
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,423A
;      FILING DATE:  06-OCT-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,654
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,573
;      FILING DATE:  JUN-07-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-105
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-944-423A-40


alignment_scores:
            Quality:  161.50              Length:   1318
              Ratio:  0.273                 Gaps:     67
 Percent Similarity:  44.917   Percent Identity: 20.865

alignment_block:
 US-09-301-906-3/rev x US-08-944-423A-40   ..

 Align seg 1/1  to: US-08-944-423A-40  from: 1  to: 1481

    3852 ATGCTGAGTCCC.........AGATGGACCAGCACCTTCGTCATAGTCAA 3812
       1 MetAlaSerProArgAlaSerArgTrpProProProLeuLeuLeuLeuLe 17

    3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCCTCTACC 3762
      17 uLeuPro.................LeuLeuLeuLeuProPro....... 25

    3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTAT 3712
      26 .....................................AlaAlaProGly 29

    3711 GATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662
      30 ThrArg.....................................AspProPr 34

    3661 TCCCTCGCCAAACAACCTCGGGAACAGC...GCACCAATTATTGGAGAAG 3615
      34 oProSerProAlaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51

    3614 CCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCAAGCAATGTCCC 3565
      51 lyLeuGluLeuGlnLeuGluArg.........ArgProGluArg.GluPr 64

    3564 TAAACCAAAGAAAAACATTAAGTGCCGTGGAAGAGCATGTGCCCAAATGC 3515
      64 oProProThrProProArgGluArgArgGlyProAla...ThrProGlyP 80

    3514 CTAACATACTGCAGCCGGAA............................. 3495
      80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96

    3494 ............................ATGGCGTATGAGGTAGAAAA 3475
      97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113

    3474 CGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACGGTCGCCCTAGTGA 3425
     113 oGluSerAsnThrGlu..............AlaHisValGluAsnIleT 125

    3424 CGGTCACCGCCGACGAAAACCCTCTTCGCACGACTAAAGATCCACCGGAT 3375
     125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141

    3374 ACTGTGCTAGAGGTTAAACCGCGTAAACCACCGGCGGCTTGTGAACTTAA 3325
     142 ValMetValGlnThrSerGlyLysSer..................... 150

    3324 CGAGCTTAGAGTTTCGTAACTACTGCCCGCGTACCTTCTAGCGTTAGCCG 3275
```

```
 150 ..................................................  150

3274 CAATTCCGCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAAGACGCC 3225
                          ||||||   :::   ||||||:::::
 151 ...................HisAlaAlaSerAspAlaProGluAsnLeu 160

3224 ACGAATATCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGTGAAGTC 3175
                        ||||||:::   :::|||      ::::::::
 161 .................ThrLeuLeuAlaGluThrAlaAspAlaArgGl 171

3174 ATTTAAGGAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGACCTTGC 3125
     :   :::   :::|||   :::::::::|||
 171 yArgSerGlySerSerSerArgThrAsnPhe..................... 181

3124 TCACAGCTACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTT...... 3081
                           :::   :::|||:::|||
 182 ...................ThrIleLeuProValGlyTyrSerLeuGlu 191

3080 ATAGCATTCTCTGTGCGCACCAGAGCATCAAGCCTTTCATCT...TCAGT 3034
     ||||||   ::::::   ::::::::::::::::|||:::|||   |||::
 192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208

3033 TAAACCGCGGTCGTAGAATAGTGAACCTGTGGAAAAGAACGACGTAATGC 2984
     :         |||   :::||||||      |||:::      ::::::
 208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225

2983 CAGAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAACCCTTAGAAACA 2934
          |||:::   :::|||:::      :::|||:::
 225 lnThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241

2933 CTCGAA..............................ACCACGGTCGCTTT 2914
     |||:::                             ||||||   |||::
 242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258

2913 CGAACCGCCAAGAAGACCTGCCCTC........................... 2889
     ::::|||         ||||||||
 258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275

2888 ............................GTACAGTGG............. 2880
                                 ::::::|||
 275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291

2879 ...ACGTTAGAAACGTTCCCCATTATGAACGCATATACAGAATCTTCAGT 2833
        |||      :::   |||:::::::::      :::   |||:::
 292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu.. 307

2832 AATTGGCAGGGGTTTTCCATCTTCTGTCTTCATTGCCTTCATGACCTTAA 2783
                       |||:::      :::         |||   :
 308 ...................SerThrGluLysLeuAsnAsnSerThrGlyL 318

2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCGC......AAGCGACTA 2739
     :::::||||||:::::::::      ||||||              :::
 318 euGlnSerSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334

2738 TTCTTCAGTTTCCTACCGAAGGTC......TCCTTCGAAATATCAAAGGC 2695
     |||   :::      |||::::::      |||:::   :::|||   ::
 335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGl 351

2694 ACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAAGTGCAAATGGTAACAA 2645
     :|||:::|||:::::::   :::         :::   :::::::::||||
 351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368

2644 CAACAGACCCAGGTACTTTACGGATGAACGCTCTGGGC............ 2607
     ||::::::      |||      :::|||:::
 368 hrSerSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384

2606 ................................ATAGACATAAGCGT 2593
                                     |||:::   |||::
 385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401

2592 CTCCAAAACGTTGAAGTTATCGTGCACCATTGCCTTAGCATAGACGGAGA 2543
     :::::::   :::   |||                        |||:::
 401 rGluAsnGluPheGlyLeu......................ThrSerL 410

2542 CTGCGTTCGAAAAACAATCGCCTACGGTCTGAAAGAAGAGCTTGAAGACC 2493
          ::::::::::::|||||||||      :::   :::|||   :::
 410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426

2492 CCGGCGAGACTCCCCTCACTGGCTTTTATAAAGTTATAGGAGCGAATACT 2443
        :::::::::   :::   :::   ::::::      |||
 427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu........ 440

2442 CTTCGCCCTATCCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393
                                :::::::::   ::::::   |||
 441 ........................ThrValSerArgSerValAlaPro. 448

2392 ACATATCATCAGGAGAAATATCAACCTTGCTGTGG.............. 2358
       :::   :::|||||||||||   |||       |||
 449 ..MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463

2357 ACAACGCGACTCCCTGTA............................... 2340
     ||||||          |||
 464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480

2339 ..........ATGACATGGGTCTTTTGCGACCTGACC.............. 2313
                :::          |||   |||   |||
 480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497

2312 ..GCCCCAACGGCATACTCAAACGTTCTATTCTCGAATGTCCCTATAGTG 2265
       :::   ||||||   ::::::      :::|||:::          :::
 497 lySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThr 513

2264 TTGAGCGCGTAGGAATAGATCCTGTCGACAAAGTCCGAATCGAGGACTAT 2215
        |||:::   |||   :::   |||:::      :::::::::
 514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529

2214 TGTCTTAAAGGACACACTAGAACCA.......................G 2189
      :::|||:::|||:::   ||||||                       |
 529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerA 546

2188 CTACAACAGGTATGAACACCAGCGATTTGCCCACGAACGACGGGACCAAC 2139
     ||:::::::      :::||||||   :::|||:::      ||||||:::
 546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561

2138 GACCTATACGTAACCTCAGT.......TCCAGGTGACTTCTGAGCCCTAC 2096
        |||   :::::::::          :::|||   |||   |||   ||
 562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578

2095 ACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTCTAGCTTA 2046
     |   ::::::      ::::::||||||
 578 rLeuThrAsnThrSerMetSerThrThr..................... 587

2045 AACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAA......AAA 2002
                                     :::|||||||      :::
 588 ..............................SerGlyGluAlaGlySer 593

2001 CCCGCTTACGTT............AGAGAAGGAATGTTCATACATCTCAC 1964
     ||||||   :::            ||||||       :::|||:::
 594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609

1963 CACACTGACCAACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914
      ::::::|||:::   :::|||   |||                  :::|
 610 .AsnValThrAspAspMetGlyLeuVal.................SerA 620

1913 GT..................................ATCGAGTTCTTC 1900
     ||                                  |||::::::
 620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636

1899 CAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACAT..... 1855
     ||||||:::   |||:::      |||:::      ::::::|||
 637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653

1854 ..............AAGAGCTATGTCGAAAACCAGCGCACCCTTCTTA 1821
                    ||||||||||||||||||||:::|||:::||| |||
 653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669
```

```
1820 TCCATAGCCTCTGCTACCTGCTGCGGGGATATA......TCGTACACATC 1777
     ||||||:::::   :::      :::||||||      |||   |||||
 670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrSe 686

1776 CACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCGAC 1727
     |   ||||||   :::::::::   :::|||:::::
 686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr............. 698

1726 TACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCA...CTG 1680
               ::::::|||:::   |||:::::::::||||||   :::
 699 ..........SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711

1679 GACACGTACGAA.........TCTCCCCCAGCTGTGGAAAGAAAGAGGAT 1639
     ::::::|||:::         :::   |||:::::|||
 712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu........... 724

1638 CTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGCATA 1589
      |||   :::|||:::   ::::::|||   |||   |||
 725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr.......... 737

1588 CATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCCCCA 1539
                                                  |||
 738 .............................................ProThr 739

1538 ATATCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAA 1489
     |||:::      :::|||
 740 IleAsnMetProAsnThr................................ 745

1488 ACAGCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGACCGAGTCGG 1439
                          :::   :::|||   |||::::::|
 746 ........................SerValValLeuAspThrAspAlaG 754

1438 AGAACTGTATCGAAAGCTCTGGAAAGAGTTCCCTCAGCTTCT........ 1397
     ||         :::|||||||:::::|||||| |||:::|||
 754 luPheValSerAspSerSerSerSerSerSer.SerSerSerSerSerSe 770

1396 ...................CCCTTTCCCGCGGCTTCAGAAACACATCGCA 1366
                        |||:::|||::::::|||::::::|||  ||
 770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis..Hi 786

1365 CCTATGTGAGGGTTTTTCGCGAATCATTGCGTCCCTTTGAGAAAACATAT 1316
     ||||         ||||||   |||:::   |||         ::::::
 786 sLeu.........PheSerSerIleLeuProSerThrArgAlaSerVal. 799

1315 CGCACAATTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAAT 1266
       |||   :::   |||      ::::::::::::   ||||||   :::
 800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815

1265 ACCAGTCTATTGGTGATACTCGAAACGACCGATGAATAGACCTCCTCTTG 1216
        |||   |||   :::      ||||||:::      ||||||:::
 816 ProSerProLeuProValSerLeuThrThrSer......ThrSerAlaPr 830

1215 ATTAAGGGCATTGGCGACGGAGGCCAAGATCTCCCCTCCCAACCGTGTCC 1166
      |||   :::      |||         |||      :::   ||||
 830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847

1165 TATCGCCAGCGACGATAGCGGTTTTTGGGACATTGGGGAGATAGACGTCT 1116
     ||      |||      :::         ::::::         |||
 847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863

1115 CCGCCCTTCCACCCAAAGGTACGCTCTTCATTGAGTGATGGCATGTGTAG 1066
           |||         ::::::||||||   :::      :::   ||
 864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880

1065 TATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAG...CCAA 1019
     |:::   :::|||   :::      ::::::::::::   |||   |||
 880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897

1018 GGTGCATCAAGAAACCTATCGAAAAG..................... 993
     .   ::::::   |||   |||:::
 897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913

 992 ...................GAAGATGAGAAGCAACGACGTCTGTAAAT 964
```

```
                              |||::::::   ::::::   |||
      914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928

      963 TTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTGGTGGTGC 914
            |||   :::   :::   |||   |||       |||       :::
      929 ....ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944

      913 ATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGGCAAGGCA 864
            |||               ||||||   :::|||:::         :::
      944 hrThr..............IleLeuLysThrSerGlnProLeuMetThr 955

      863 ACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATC........ 822
          |||      ||||||::::::   :::      |||   :::
      956 Thr......ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970

      821 .GTAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGA 773
           :::|||:::      |||   ||||||:::         ||||||||||
      970 oIleAlaValGlnThrThrAlaGlyLysGln.........LeuSerLeuT 984

      772 CGTAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAG......ATTTTA 729
          ||           |||:::  :::   |||::::::      |||
      984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000

      728 GCCCAAACGAACGTGTTCTTTACATCAATAGGAACATTAAGGGCCCTGTT 679
          ::::::   |||      :::::::::      |||            ||
     1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017

      678 AAACAATTTCCCT......TTCCGCAAGATGACAGCGTCGCTCTCA.... 639
          |::::::   |||      ::::::||||||::::::|||:::
     1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034

      638 ...........TCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAG 600
                    ||||||              :::||| :::   :::
     1034 hrAlaGluTyrSerPro............AlaSerArg.SerLeuGlyTh 1046

      599 GCGGCCTTCTCCTCCAACGACTGTTGTGGGACT.......GATTGCGCTG 557
              |||||||||   ||||||||||||:::          |||   |||
     1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063

      556 GCTCCAGTGCTTCCTGTTTTTTAGGAAGTGGTACGTGGACGCCAACAGCT 507
            :::||||||:::      :::   |||:::|||   :::||||||:::
     1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079

      506 GTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGCTTCTGC 462
          ::::::   :::|||:::   :::::::|||   ::::::   |||
     1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094
```

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-944-496-40

```
seq_documentation_block:
; Sequence 40, Application US/08944496
; Patent No. 6124433
;  GENERAL INFORMATION:
;    APPLICANT:  FALB, DEAN A
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
;    NUMBER OF SEQUENCES:  54
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/944,496
;      FILING DATE:  06-OCT-1997
;      CLASSIFICATION:  514
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,654
;      FILING DATE:  09-FEB-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/485,573
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,844
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CORUZZI, LAURA A
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  7853-104
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 790-9090
;      TELEFAX:  (212) 869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1481 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-944-496-40


alignment_scores:
            Quality:  161.50              Length:   1318
              Ratio:  0.273                Gaps:     67
 Percent Similarity:  44.917   Percent Identity: 20.865

alignment_block:
 US-09-301-906-3/rev x US-08-944-496-40   ..

 Align seq 1/1  to: US-08-944-496-40  from: 1  to: 1481

    3852 ATGCTGAGTCCC.........AGATGGACCAGCACCTTCGTCATAGTCAA 3812
           |||   ||||||         ||||||          ::::::::::::
       1 MetAlaSerProArgAlaSerArgTrpProProProLeuLeuLeuLeuLe 17

    3811 AGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCCCTCTACC 3762
          :::|||                  ::::::   :::||||||
      17 uLeuPro.................LeuLeuLeuLeuProPro....... 25

    3761 GGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTAT 3712
                                                  ||||||
      26 .....................................AlaAlaProGly 29

    3711 GATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCTGACACTCT 3662
            :::                                    |||
      30 ThrArg.................................AspProPr 34

    3661 TCCCTCGCCAAACAACCTCGGGAACAGC...GCACCAATTATTGGAGAAG 3615
          |||||||||   :::   :::   |||   ||||||:::   |||   :
      34 oProSerProAlaArgArgAlaLeuSerLeuAlaProLeuAlaGlyAlaG 51

    3614 CCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCAAGCAATGTCCC 3565
         ::   :::|||         |||         ||||||::::::    ||
      51 lyLeuGluLeuGlnLeuGluArg.........ArgProGluArg.GluPr 64

    3564 TAAACCAAAGAAAAACATTAAGTGCCGTGGAAGAGCATGTGCCCAAATGC 3515
         |   |||            :::   ||||||   |||   :::      |
      64 oProProThrProProArgGluArgArgGlyProAla...ThrProGlyP 80

    3514 CTAACATACTGCAGCCGGAA.............................. 3495
         ||:::       ||||||
      80 roSerTyrArgAlaProGluProGlyAlaAlaThrGlnArgGlyProSer 96

    3494 ............................ATGGCGTATGAGGTAGAAAA 3475
                                     |||::::::
      97 GlyArgAlaProArgGlyGlySerAlaAspAlaAlaTrpLysHisTrpPr 113

    .       .

    3474 CGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACGGTCGCCCTAGTGA 3425
          :::||||||::::::               |||   |||      :::|
     113 oGluSerAsnThrGlu.............AlaHisValGluAsnIleT 125

    3424 CGGTCACCGCCGACGAAAACCCTCTTCGCACGACTAAAGATCCACCGGAT 3375
         ||         :::::::::   :::   |||:::::::::
     125 hrPheTyrGlnAsnGlnGluAspPheSerThrValSerSerLysGluGly 141

    3374 ACTGTGCTAGAGGTTAAACCGCGTAAACCACCGGCGGCTTGTGAACTTAA 3325
         ::::::::::::::::::   ::::::
     142 ValMetValGlnThrSerGlyLysSer..................... 150

    3324 CGAGCTTAGAGTTTCGTAACTACTGCCCGCGTACCTTCTAGCGTTAGCCG 3275

     150 ................................................. 150

    3274 CAATTCCGCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAAGACGCC 3225
                           ||||||   :::   ||||||::::::
     151 ...................HisAlaAlaSerAspAlaProGluAsnLeu 160

    3224 ACGAATATCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGTGAAGTC 3175
                         ||||||:::   :::|||       ::::::::
     161 ..................ThrLeuLeuAlaGluThrAlaAspAlaArgGl 171

    3174 ATTTAAGGAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGACCTTGC 3125
         :   :::   :::|||   :::::::::|||
     171 yArgSerGlySerSerSerArgThrAsnPhe................... 181

    3124 TCACAGCTACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTT...... 3081
                           :::   :::|||:::|||
     182 ...................ThrIleLeuProValGlyTyrSerLeuGlu 191

    3080 ATAGCATTCTCTGTGCGCACCAGAGCATCAAGCCTTTCATCT...TCAGT 3034
         ||||||   ::::::   ::::::::::::::|||:::|||   |||::
     192 IleAlaThrAlaLeuThrSerGlnSerGlyAsnLeuAlaSerGluSerLe 208

    3033 TAAACCGCGGTCGTAGAATAGTGAACCTGTGGAAAAGAACGACGTAATGC 2984
         :         |||   :::||||||      |||:::      ::::::
     208 uHisLeuProSerSerSerSerGluPheAspGluArgIleAlaAlaPheG 225

    2983 CAGAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAACCCTTAGAAACA 2934
              |||:::   :::|||:::      :::|||:::
     225 lnThrLysSerGlyThrAlaSerGluMetGlyThrGluArgAlaMetGly 241

    2933 CTCGAA............................ACCACGGTCGCTTT 2914
         |||:::                           ||||||   |||::
     242 LeuSerGluGluTrpThrValHisSerGlnGluAlaThrThrSerAlaTr 258

    2913 CGAACCGCCAAGAAGACCTGCCCTC........................ 2889
         ::::|||         |||||||||
     258 pSerProSerPheLeuProAlaLeuGluMetGlyGluLeuThrThrProS 275

    2888 ............................GTACAGTGG............ 2880
                                     ::::::|||
     275 erArgLysArgAsnSerSerGlyProAspLeuSerTrpLeuHisPheTyr 291

    2879 ...ACGTTAGAAACGTTCCCCATTATGAACGCATATACAGAATCTTCAGT 2833
            |||      :::   |||:::::::::      :::   |||:::
     292 ArgThrAlaAlaSerSerProLeuLeuAspLeuSerSerProSerGlu.. 307

    2832 AATTGGCAGGGGTTTTCCATCTTCTGTCTTCATTGCCTTCATGACCTTAA 2783
                           |||:::      :::         |||   :
     308 ...................SerThrGluLysLeuAsnAsnSerThrGlyL 318

    2782 TTGAATCTTCCACGATAGCCCTAGAGAAGACGCGC......AAGCGACTA 2739
         :::::||||||:::::::::      ||||||               :::
     318 euGlnSerSerSerValSerGlnThrLysThrMetHisValAlaThrVal 334

    2738 TTCTTCAGTTTCCTACCGAAGGTC......TCCTTCGAAATATCAAAGGC 2695
         |||   :::      |||::::::      |||:::   :::|||   ::
     335 PheThrAspGlyGlyProArgThrLeuArgSerLeuThrValSerLeuGl 351

    2694 ACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAAGTGCAAATGGTAACAA 2645
```

```
 351 yProValSerLysThrGluGlyPheProLysAspSerArgIleAlaThrT 368

2644 CAACAGACCCAGGTACTTTACGGATGAACGCTCTGGGC..... 2607
 368 hrSerSerSerValLeuLeuSerProSerAlaValGluSerArgArgAsn 384

2606 .....ATAGACATAAGCGT 2593
 385 SerArgValThrGlyAsnProGlyAspGluGluPheIleGluProSerTh 401

2592 CTCCAAAACGTTGAAGTTATCGTGCACCATTGCCTTAGCATAGACGGAGA 2543
 401 rGluAsnGluPheGlyLeu.....ThrSerL 410

2542 CTGCGTTCGAAAAACAATCGCCTACGGTCTGAAAGAAGAGCTTGAAGACC 2493
 410 euArgTrpGlnAsnAspSerProThrPheGlyGluHisGlnLeuAlaSer 426

2492 CCGGCGAGACTCCCCTCACTGGCTTTTATAAAGTTATAGGAGCGAATACT 2443
 427 SerSerGluValGlnAsnGlySerProMetSerGlnThrGlu..... 440

2442 CTTCGCCCTATCCTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCC 2393
 441 .....ThrValSerArgSerValAlaPro. 448

2392 ACATATCATCAGGAGAAATATCAACCTTGCTGTGG..... 2358
 449 ..MetArgGlyGlyGluIle...ThrAlaHisTrpLeuLeuThrAsnSer 463

2357 ACAACGCGACTCCCTGTA..... 2340
 464 ThrThrSerAlaAspValThrGlySerSerAlaSerTyrProGluGlyVa 480

2339 .....ATGACATGGGTCTTTTGCGACCTGACC..... 2313
 480 lAsnAlaSerValLeuThrGlnPheSerAspSerThrValGlnSerGlyG 497

2312 ..GCCCCAACGGCATACTCAAACGTTCTATTCTCGAATGTCCCTATAGTG 2265
 497 lySerHisThrAlaLeuGlyAspArgSerTyrSerGluSerSerSerThr 513

2264 TTGAGCGCGTAGGAATAGATCCTGTCGACAAAGTCCGAATCGAGGACTAT 2215
 514 SerSerSer...GluSerLeuAsnSerSerAlaProArgGlyGluArgSe 529

2214 TGTCTTAAAGGACACACTAGAACCA.....G 2189
 529 rThrLeuGluAspSerArgGluProGlyGlnAlaLeuGlyAspSerSerA 546

2188 CTACAACAGGTATGAACACCAGCGATTTGCCCACGAACGACGGGACCAAC 2139
 546 laAsnAlaGluAspArgThrSerGlyValProSerLeu...GlyThrHis 561

2138 GACCTATACGTAACCTCAGT.....TCCAGGTGACTTCTGAGCCCTAC 2096
 562 ThrLeuAlaThrValThrGlyAsnGlyGluArgThrLeuArgSerValTh 578

2095 ACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTCTAGCTTA 2046
 578 rLeuThrAsnThrSerMetSerThrThr..... 587

2045 AACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAA.....AAA 2002
 588 .....SerGlyGluAlaGlySer 593

2001 CCCGCTTACGTT.....AGAGAAGGAATGTTCATACATCTCAC 1964
 594 ProAlaAlaAlaMetProGlnGluThrGluGlyAlaSerLeuHisVal.. 609

1963 CACACTGACCAACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAAC 1914
```

```
 610 .AsnValThrAspAspMetGlyLeuVal.....SerA 620

1913 GT.....ATCGAGTTCTTC 1900
 620 rgSerLeuAlaAlaSerSerAlaLeuGlyValAlaGlyIleSerTyrGly 636

1899 CAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACAT..... 1855
 637 GlnValArgGlyThrAlaIleGluGlnArgThrSerSerAspHisThrAs 653

1854 .....AAGAGCTATGTCGAAAACCAGCGCACCCTTCTTA 1821
 653 pHisThrTyrLeuSerSerThrPheThrLysGlyGluArgAlaLeu.Leu 669

1820 TCCATAGCCTCTGCTACCTGCTGCGGGGATATA.....TCGTACACATC 1777
 670 SerIleThrAspAsnSerSerSerSerAspIleValGluSerSerThrSe 686

1776 CACCATAAAACCCGCATCGGCTCTAGAATCGCAGTTCTGCGATTCTCGAC 1727
 686 rTyrIleLysIleSerAsnSerSerHisSerGluTyr..... 698

1726 TACAGTAAGACACAGACTTTGAAGCCGCTTCAGTTAGAAGGTCA...CTG 1680
 699 .:.....SerSerPheSerHisAlaGlnThrGluArgSerAsnIle 711

1679 GACACGTACGAA.....TCTCCCCCAGCTGTGGAAAGAAAGAGGAT 1639
 712 SerSerTyrAspGlyGluTyrAlaGlnProSerThrGlu..... 724

1638 CTCATTGATTCTCCGCTTCACATCTTTAACGTCTAGGACTGGATTGCATA 1589
 725 .SerProValLeuHisThrSerAsnLeuProSerTyrThr..... 737

1588 CATGACAGTTCACATGGCCAGCTTTGACATGATACGTGAAGCTCCCCCCA 1539
 738 .....ProThr 739

1538 ATATCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAA 1489
 740 IleAsnMetProAsnThr..... 745

1488 ACAGCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGACCGAGTCGG 1439
 746 .....SerValValLeuAspThrAspAlaG 754

1438 AGAACTGTATCGAAAGCTCTGGAAAGAGTTCCCTCAGCTTCT..... 1397
 754 luPheValSerAspSerSerSerSerSerSer.SerSerSerSerSerSe 770

1396 .....CCCTTTCCCGCGGCTTCAGAAACACATCGCA 1366
 770 rSerSerGlyProProLeuProLeuProSerValSerGlnSerHis..Hi 786

1365 CCTATGTGAGGGTTTTTCGCGAATCATTGCGTCCCTTTGAGAAAACATAT 1316
 786 sLeu.....PheSerSerIleLeuProSerThrArgAlaSerVal. 799

1315 CGCACAATTTCGTGTCCAAATGGGAAAGCAATGCCGATTGGTCCCTTAAT 1266
 800 ..HisLeuLeuLysSerThrSerAspAlaSerThrProTrpSerSerSer 815

1265 ACCAGTCTATTGGTGATACTCGAAACGACCGATGAATAGACCTCCTCTTG 1216
 816 ProSerProLeuProValSerLeuThrThrSer.....ThrSerAlaPr 830

1215 ATTAAGGGCATTGGCGACGGAGGCCAAGATCTCCCCTCCCAACCGTGTCC 1166
 830 oLeuSerValSerGlnThrThrLeuProGlnSerSerSerThrProValL 847

1165 TATCGCCAGCGACGATAGCGGTTTTTGGGACATTGGGGAGATAGACGTCT 1116
 847 euProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet 863
```

```
 1115 CCGCCCTTCCACCCAAAGGTACGCTCTTCATTGAGTGATGGCATGTGTAG 1066
          |||         ::::::||||||   :::      :::   ||
  864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAspLeuLysSe 880

 1065 TATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAG...CCAA 1019
       |:::   :::|||   :::     ::::::::::::   |||   |||
  880 rGlnSerThrProHisGlnGluLysValIleThrGluSerLysSerProS 897

 1018 GGTGCATCAAGAAACCTATCGAAAAG..................... 993
       ::::::   |||   |||:::
  897 erLeuValSerLeuProThrGluSerThrLysAlaValThrThrAsnSer 913

  992 ..................GAAGATGAGAAGCAACGACGTCTGTAAAT 964
                         |||::::::   ::::::   |||
  914 ProLeuProProSerLeuThrGluSerSerThrGluGlnThrLeu..... 928

  963 TTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTGGTGGTGC 914
        |||   :::   :::   |||   |||      |||      :::
  929 ....ProAlaThrSerThrAsnLeuAlaGlnMetSerProThrPheThrT 944

  913 ATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGGCAAGGCA 864
        |||              ||||||   :::|||:::         :::
  944 hrThr..............IleLeuLysThrSerGlnProLeuMetThr 955

  863 ACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATC........ 822
      |||      ||||||::::::   :::      |||   :::
  956 Thr......ProGlyThrLeuSerSerThrAlaSerLeuValThrGlyPr 970

  821 .GTAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGA 773
       :::|||:::      |||   ||||||:::         ||||||||||
  970 oIleAlaValGlnThrThrAlaGlyLysGln.........LeuSerLeuT 984

  772 CGTAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAG......ATTTTA 729
      ||           |||:::   :::   |||:::::      |||
  984 hrHisProGluIleLeuValProGlnIleSerThrGluGlyGlyIleSer 1000

  728 GCCCAAACGAACGTGTTCTTTACATCAATAGGAACATTAAGGGCCCTGTT 679
      ::::::   |||      :::::::::      |||            ||
 1001 ThrGluArgAsnArgValIleValAspAlaThrThrGlyLeuIleProLe 1017

  678 AAACAATTTCCCT......TTCCGCAAGATGACAGCGTCGCTCTCA.... 639
      |:::::   |||        ::::::||||||::::::|||:::
 1017 uThrSerValProThrSerAlaLysGluMetThrThrLysLeuGlyValT 1034

  638 ...........TCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAG 600
                 ||||||            :::|||  :::   :::
 1034 hrAlaGluTyrSerPro............AlaSerArg.SerLeuGlyTh 1046

  599 GCGGCCTTCTCCTCCAACGACTGTTGTGGGACT.......GATTGCGCTG 557
        |||||||||   ||||||||||||:::         |||   |||
 1046 rSerProSerProGlnThrThrValValSerThrAlaGluAspLeuAlaP 1063

  556 GCTCCAGTGCTTCCTGTTTTTTAGGAAGTGGTACGTGGACGCCAACAGCT 507
       :::||||||:::      :::   |||:::|||   :::||||||:::
 1063 roLysSerAlaThrPheAlaValGlnSerSerThrGlnSerProThrThr 1079

  506 GTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGCTTCTGC 462
      ::::::   :::|||:::   ::::::|||   ::::::   |||
 1080 LeuSerSerSerAlaSerValAsnSerCysAlaValAsnProCys 1094

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-682-517-15

seq_documentation_block:
; Sequence 15, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-15


alignment_scores:
            Quality:  150.50            Length:    927
              Ratio:  0.358               Gaps:     45
 Percent Similarity:  45.307   Percent Identity: 19.957

alignment_block:
 US-09-301-906-3 x US-08-682-517-15   ..

 Align seg 1/1  to: US-08-682-517-15  from: 1  to: 1222

 3784 GATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGG 3833
      |||   :::|||   :::   |||      |||      ||||||
  164 AspPheAlaLeuValPheLysArgThrValAspLysValGluGlyGluTh 180

 3834 TCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGAT 3883
       |||           :::||||||:::                :::::::
  180 rPro...........GluGluAlaAlaPheValLysAlaIleAsnAsnT 193

 3884 CCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTG 3933
      ::|||   :::|||:::      :::   ||||||||||||:::   :::
  193 hrThrValGluValThrPheGluGluGluValThrAsnValGlnAlaLeu 209

 3934 GGGATTAGCACCAACGGG...GAAGTTACTGGTGAAGAAGAGACCCATTC 3980
      :::::::::   :::|||   |||:::   :::   :::      :::::
  210 AsnPheLysIleGluGlyLeuGluIleLysAsnAlaSerValLysGlnTh 226

 3981 ACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTG 4030
      :   ::::::|||          :::|||   :::::::|||   :::
  226 rAsnLysLysValValValLeuThrThrGluAlaGlnThrAlaAspLysG 243

 4031 CAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGT......TCTGGTACC 4074
           |||         :::|||:::      |||      :::|||:::
  243 luTyrValLeuThrLeuAspGlyGluThrIleGlyGlyPheLysGlyVal 259

 4075 GCTGACGCTATG...GCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGA 4121
      |||   ::::::   :::   ||||||   |||::::::::::|||
  260 AlaAlaValValProThrLysValGluLeuValSerSerAlaVal..... 274

 4122 TGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGG 4171
                        :::|||:::   ::::::||||||:::|||:
  275 ..................GlnGlyLysLeuGlyGlnGluValLysValG 285

 4172 ACGGCAAA..................................... 4179
      :::::|||
  285 lnAlaLysValThrValAlaGluGlyGlnSerLysAlaGlyIleProVal 301

 4180 ............................GGGTTGCTCCCAGAAAGCGT 4199
                                  |||::::::|||   :::::
  302 ThrPheThrValProGlyAsnAsnAsnAspGlyValValPro...ThrLe 317

 4200 CGTCGGTGAGGCGCCGACACAAGAAAGGGGA.................. 4230
      ::::|||||||||   |||:::|||:::|||
  317 uThrGlyGluAlaLeuThrAsnGluGluGlyIleAlaThrTyrSerTyrT 334

 4231 ..AGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGAC 4278
        |||      :::|||:::         ||||||         ||||||
  334 hrArgTyrLysGluGlyThrAspGluValThrAlaTyrAlaThrGlyAsp 350
```


4279 AGG 4281
351 ArgSerLysPheSerLeuGlyTyrValPheTrpGlyValAspThrIleLe 367
4282 GAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTG 4321
367 uSerValGluGluValThrThrGlyAlaSerValAsnAsnGlyAlaAsnL 384
4322 ATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAA...GAAGTG 4368
384 ysThrTyrLysValThrTyrLysAsnProLysThrGlyLysProGluAla 400
4369 AAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGA 4418
401 AsnLysThrPheAsnValGlyPheValGluAsnMetAsn...ValThrSe 416
4419 TGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACA 4468
416 rAspLysValAla...AsnAlaThrValAsnGlyValLysAlaLeuGlnL 432
4469 AGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAG 4518
432 euSerAsnGlyThrAla......LeuAspAlaAlaGlnIleThrThrAsp 446
4519 CAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACA 4568
447 SerLysGlyGluAlaThrPheThrValSerGlyThrAsnAlaAlaValTh 463
4569 GCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACA 4618
463 rProValValTyr.....AspLeuHisSerThrA 473
4619 ACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
473 snAsnSerThrSerAsnLysLysTyrSerAlaSerAlaLeuGlnThrThr 489
4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTA 4718
490 AlaSerLysValThrPheAlaAlaLeuGlnAlaGluTy 502
4719 TACCTAT.....AATGGTGGTAGCCATGTTTCATCAG 4750
502 rThrIleGluLeuThrArgAlaAspAsnAlaGlyGlu..... 514
4751 GGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGC 4800
515 ValAlaAlaIleGlyAlaThrAsn 522
4801 GTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGA..... 4842
523 GlyArgGluTyrLysValIleValLysAspLysAlaGlyAsnLeuAlaLy 539
4843 GGCGTACCATTCCACGCTGATGACGAGGAGTGC.... 4875
539 sAsnGluIleValAsnValAlaPheAsnGluAspLysAspArgValIleS 556
4876 TATCCATCAGATAACCCTATCTTGACGGTC 4905
556 erThrValThrAsnAlaLysPheValAspThrAspProAspThrAlaVal 572
4906 AATCTCGTGGGG...AAGGCA...AACTTCTCGACTAAGTGCAGG...AA 4946
573 TyrPheThrGlyAspLysAlaLysGlnIleSerValLysThrAsnAspLy 589
4947 GGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTA 4996
589 sGlyGluAlaThrPheValIleGlySerAspThrValAsnAspTyrAlaT 606
4997 TGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGAC 5046
606 hrProIleAlaTrp.....IleAspIleAsnThrSerAsp 617
5047 GAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGT 5096
618 AlaLysGln..... 620
5097 AGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAG 5146
621 GlyAspLeuAspGluGlyGluProL 629
5147 GTGTTCCAAAC.....CAGCAACCACAGAGCCAAGGTGCT 5181
629 ysAlaValAlaProIleSerTyrPheGlnAlaProTyrLeuAspGlySer 645
5182 ACCAGAACAATCACACCAAA 5201
646 AlaIleLysAlaTyrLysLysSerAspLeuAsnLysAlaValThrLysPh 662
5202 ATCGGGGGGCAAGGCT.....CTATCTGAGGGAAGTG 5233
662 eAspGlySerGluThrAlaValPheAlaAlaGluLeuValAsnGlnSerG 679
5234 GTAGGGAAGTCAAGGGG.....AGGTCGACATACTCGATATGG 5271
679 lyLysLysValThrGlyThrSerIleLysLysAlaThrTyrThrIle... 694
5272 TGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCC 5321
695 TyrAsnThrGlyAlaAsnAspIleLysValAspAsnGl 707
5322 AGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTT. 5370
707 nValIleSerProAsnArgSerTyrThr...ValThrTyrGluAlaThrL 723
5371 AAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTAT 5415
723 euSerSerThrGlyThr..... 728
5416 CTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAA 5465
729 ValIleThrProAlaLysAs 735
5466 TATTGCCGTCACTACCGCCGAAGGT.....GTTCTGAAAGTACCTA 5506
735 nLeuGluValThrSerValAspGlyLysThrThrAlaValLysVal.... 750
5507 ATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACA 5556
751 IleAlaThrGlyIleAlaValAsnThrAspGlyLys 762
5557 GATCTCATTTTT.....CATTCAACACAAGACGGCTTGCGTGTGAG 5597
763 AspTyrAlaPheThrAlaLysGluAlaThrAlaThrPheThrAlaThrAs 779
5598 AGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCA 5647
779 nGluValProAsn.....SerTyrThrGlyValAlaThrGlnPheA 793
5648 AAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGAC 5697
793 snThrAlaAspSer.....GlySerAsn..... 800
5698 GATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGT 5747
801 SerAsnSerIleTrpPheAlaGlyLysAsnProVa 812
5748 CGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTC..... 5787
812 lLysTyrAlaGlyValSerGlyLysThrTyrLysTyrPheGlyAlaAsnG 829
5788 GAATATAAGTGCTATAATGCT 5808
829 lyAsnGluValPheGlyGluAlaAlaTrpGluAlaLeuLeuThrGlnTyr 845
5809 CCCCCAGGTGGCGGTAAGACGACG.....AC 5834

```
     846 AlaThrGluGlyGlnLysValThrIleSerTyrAsnValAspGlyAspTh 862

    5835 GTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCC..... 5874
         |:::::::   :::   :::::::::   ||||||::::::
     862 rValThrPheLysValIleSerAlaValAsnSerSerThrGluAlaIleL 879

    5875 .....ACC 5877
                 |||
     879 ysProValAlaProThrThrProAlaAlaProThrThrGlyAlaLeuThr 895

    5878 ATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAA 5927
         :::|||      :::|||:::      |||::::::   |||::::::
     896 LeuThrProAlaAlaGlyGlyLeuValAspLeuThrThrAlaThrAsnTh 912

    5928 GAGAGATCCGAATTTG......GAAGGTCTCAAC...AGTGCTACCACAG 5968
                  :::|||           ||||||   ||||||||||||||
     912 rLeuGlyIleSerLeuAlaAspAlaAspLeuAsnValSerAlaThrThrV 929

    5969 TTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTT 6018
         ||::::::   :::|||:::
     929 alAspThrAlaThrValSer..... 935

    6019 TTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGT 6068
         |||   |||::::::::::            :::   |||:::|||   ::
     936 LeuLysAspSerAlaAsn.....AsnSerLeuSerLeuThrLe 948

    6069 CTTCGCAACCGGCGCGTCGGAAGGCCTCTTT 6099
         :      |||||||||:::   |||:::|||
     948 uValGluThrGlyAlaAsnThrGlyValPhe 958

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-682-517-9

seq_documentation_block:
; Sequence 9, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-9


alignment_scores:
            Quality:  150.50            Length:    927
              Ratio:  0.358              Gaps:     45
 Percent Similarity:  45.307   Percent Identity: 19.957

alignment_block:
 US-09-301-906-3 x US-08-682-517-9   ..

 Align seq 1/1  to: US-08-682-517-9  from: 1  to: 1252

    3784 GATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGG 3833
         |||   ::::|||   :::   |||      |||      ||||||
      194 AspPheAlaLeuValPheLysArgThrValAspLysValGluGlyGluTh 210

    3834 TCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGAT 3883
          |||           :::||||||:::              :::::::

     210 rPro.....GluGluAlaAlaPheValLysAlaIleAsnAsnT 223

    3884 CCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTG 3933
         ::|||   :::|||:::      :::   ||||||||||||:::   :::
     223 hrThrValGluValThrPheGluGluValThrAsnValGlnAlaLeu 239

    3934 GGGATTAGCACCAACGGG...GAAGTTACTGGTGAAGAAGAGACCCATTC 3980
         :::::::::   :::|||   |||:::   :::   :::      :::::
     240 AsnPheLysIleGluGlyLeuGluIleLysAsnAlaSerValLysGlnTh 256

    3981 ACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTG 4030
         :   :::::::|||            :::|||   :::::::|||   :::
     256 rAsnLysLysValValValLeuThrThrGluAlaGlnThrAlaAspLysG 273

    4031 CAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGT......TCTGGTACC 4074
               |||         :::|||:::      |||      :::|||:::
     273 luTyrValLeuThrLeuAspGlyGluThrIleGlyGlyPheLysGlyVal 289

    4075 GCTGACGCTATG...GCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGA 4121
         |||   ::::::   :::   ||||||   |||:::::::::|||
     290 AlaAlaValValProThrLysValGluLeuValSerSerAlaVal..... 304

    4122 TGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGG 4171
                           :::|||:::   ::::::||||||:::|||:
     305 ..................GlnGlyLysLeuGlyGlnGluValLysValG 315

    4172 ACGGCAAA..................................... 4179
         :::::|||
     315 lnAlaLysValThrValAlaGluGlyGlnSerLysAlaGlyIleProVal 331

    4180 ..........................GGGTTGCTCCCAGAAAGCGT 4199
                                    |||::::::|||   :::::
     332 ThrPheThrValProGlyAsnAsnAsnAspGlyValValPro...ThrLe 347

    4200 CGTCGGTGAGGCGCCGACACAAGAAAGGGGA.................. 4230
         ::::||||||||||   |||:::|||:::|||
     347 uThrGlyGluAlaLeuThrAsnGluGluGlyIleAlaThrTyrSerTyrT 364

    4231 ..AGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGAC 4278
           |||      :::|||:::         ||||||         ||||||
     364 hrArgTyrLysGluGlyThrAspGluValThrAlaTyrAlaThrGlyAsp 380

    4279 AGG..... 4281
         |||
     381 ArgSerLysPheSerLeuGlyTyrValPheTrpGlyValAspThrIleLe 397

    4282 ..........GAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTG 4321
                  |||   :::   ::::::   |||::::::
     397 uSerValGluGluValThrThrGlyAlaSerValAsnAsnGlyAlaAsnL 414

    4322 ATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAA...GAAGTG 4368
               |||||||||   ::::::      :::   |||   |||:::
     414 ysThrTyrLysValThrTyrLysAsnProLysThrGlyLysProGluAla 430

    4369 AAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGA 4418
         ::::::         ::::::   |||::::::   :::   ::::::::
     431 AsnLysThrPheAsnValGlyPheValGluAsnMetAsn...ValThrSe 446

    4419 TGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACA 4468
         ::::|||   |||   ::::::|||::::::|||||||||   ::::::
     446 rAspLysValAla...AsnAlaThrValAsnGlyValLysAlaLeuGlnL 462

    4469 AGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAG 4518
           :::::::::::|||         :::      |||:::   ::::::
     462 euSerAsnGlyThrAla......LeuAspAlaAlaGlnIleThrThrAsp 476

    4519 CAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACA 4568
         :::      :::   ::::::|||::::::      :::|||:::
     477 SerLysGlyGluAlaThrPheThrValSerGlyThrAsnAlaAlaValTh 493

    4569 GCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACA 4618
               ::::::                       :::|||         |
     493 rProValValTyr.....................AspLeuHisSerThrA 503
```

```
4619 ACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGAT 4668
     ||:::::|||         ::::::|||   :::         :::
 503 snAsnSerThrSerAsnLysLysTyrSerAlaSerAlaLeuGlnThrThr 519

4669 AACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTA 4718
          |||   |||   |||:::   |||          ||
 520 .....AlaSerLysValThrPheAlaAlaLeuGlnAlaGluTy 532

4719 TACCTAT.................AATGGTGGTAGCCATGTTTCATCAG 4750
     ||||                    |||:::|||:::
 532 rThrIleGluLeuThrArgAlaAspAsnAlaGlyGlu..... 544

4751 GGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGC 4800
                    ::::::||||||        :::
 545 .......................ValAlaAlaIleGlyAlaThrAsn 552

4801 GTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGA..... 4842
          ::::::   ::::::   |||   |||   |||:::
 553 GlyArgGluTyrLysValIleValLysAspLysAlaGlyAsnLeuAlaLy 569

4843 .............GGCGTACCATTCCACGCTGATGACGAGGAGTGC.... 4875
                 :::|||   |||:::   |||   ::::::
 569 sAsnGluIleValAsnValAlaPheAsnGluAspLysAspArgValIleS 586

4876 ...................TATCCATCAGATAACCCTATCTTGACGGTC 4905
                        :::   :::   :::|||      :::|||
 586 erThrValThrAsnAlaLysPheValAspThrAspProAspThrAlaVal 602

4906 AATCTCGTGGGG...AAGGCA...AACTTCTCGACTAAGTGCAGG...AA 4946
        ::::::|||   ||||||   ::::::|||:::|||   :::   ||
 603 TyrPheThrGlyAspLysAlaLysGlnIleSerValLysThrAsnAspLy 619

4947 GGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTA 4996
     ||||      ::::::||||||:::      :::   :::   :::
 619 sGlyGluAlaThrPheValIleGlySerAspThrValAsnAspTyrAlaT 636

4997 TGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGAC 5046
       |||   ::::::                 ::::::|||:::   |||
 636 hrProIleAlaTrp..............IleAspIleAsnThrSerAsp 647

5047 GAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGT 5096
          :::
 648 AlaLysGln..................................... 650

5097 AGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAG 5146
                           |||      ||||||   :::|||
 651 .......................GlyAspLeuAspGluGlyGluProL 659

5147 GTGTTCCAAAC..............CAGCAACCACAGAGCCAAGGTGCT 5181
       :::                    |||   |||      :::|||:::
 659 ysAlaValAlaProIleSerTyrPheGlnAlaProTyrLeuAspGlySer 675

5182 ..........................ACCAGAACAATCACACCAAA 5201
                                :::::::::::::|||
 676 AlaIleLysAlaTyrLysLysSerAspLeuAsnLysAlaValThrLysPh 692

5202 ATCGGGGGGCAAGGCT.................CTATCTGAGGGAAGTG 5233
      :::|||:::::::::                  |||   :::   ||||
 692 eAspGlySerGluThrAlaValPheAlaAlaGluLeuValAsnGlnSerG 709

5234 GTAGGGAAGTCAAGGGG............AGGTCGACATACTCGATATGG 5271
     ||:::::::|||   |||          :::::|||||||:::|||
 709 lyLysLysValThrGlyThrSerIleLysLysAlaThrTyrThrIle... 724

5272 TGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCC 5321
               |||         ::::::   :::::::::||||||
 725 ............TyrAsnThrGlyAlaAsnAspIleLysValAspAsnGl 737

5322 AGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTT. 5370
      |||::::::   :::   :::||||||   :::|||:::|||::::::
 737 nValIleSerProAsnArgSerTyrThr...ValThrTyrGluAlaThrL 753

5371 .....AAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTAT 5415
          :::|||||||||
 753 euSerSerThrGlyThr.............................. 758

5416 CTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAA 5465
                                        ::::::      |||:::||
 759 ............................ValIleThrProAlaLysAs 765

5466 TATTGCCGTCACTACCGCCGAAGGT.........GTTCTGAAAGTACCTA 5506
     |:::   ||||||:::::::::|||         ::::::||||||
 765 nLeuGluValThrSerValAspGlyLysThrThrAlaValLysVal.... 780

5507 ATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACA 5556
                               |||:::   |||      :::|||
 781 ..............IleAlaThrGlyIleAlaValAsnThrAspGlyLys 792

5557 GATCTCATTTTT.........CATTCAACACAAGACGGCTTGCGTGTGAG 5597
     |||      |||          ::::::|||                :::::
 793 AspTyrAlaPheThrAlaLysGluAlaThrAlaThrPheThrAlaThrAs 809

5598 AGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCA 5647
     ::::::|||            :::      |||:::   |||:::   :
 809 nGluValProAsn.........SerTyrThrGlyValAlaThrGlnPheA 823

5648 AAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGAC 5697
     ::   :::|||:::           |||:::|||
 823 snThrAlaAspSer............GlySerAsn.............. 830

5698 GATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGT 5747
                        ::::::|||   :::   |||         ||
 831 ..............SerAsnSerIleTrpPheAlaGlyLysAsnProVa 842

5748 CGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTC.......... 5787
     |      :::|||:::   |||:::::::::|||   |||
 842 lLysTyrAlaGlyValSerGlyLysThrTyrLysTyrPheGlyAlaAsnG 859

5788 ............................GAATATAAGTGCTATAATGCT 5808
                                ||||          :::
 859 lyAsnGluValPheGlyGluAlaAlaTrpGluAlaLeuLeuThrGlnTyr 875

5809 CCCCCAGGTGGCGGTAAGACGACG.......................AC 5834
           |||   |||:::|||                          ||
 876 AlaThrGluGlyGlnLysValThrIleSerTyrAsnValAspGlyAspTh 892

5835 GTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCC.......... 5874
     |:::::   :::   :::::::::   |||||:::::::
 892 rValThrPheLysValIleSerAlaValAsnSerSerThrGluAlaIleL 909

5875 ..............................................ACC 5877
                                                    |||
 909 ysProValAlaProThrThrProAlaAlaProThrThrGlyAlaLeuThr 925

5878 ATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAA 5927
     :::|||      :::|||:::      |||::::::   |||::::::
 926 LeuThrProAlaAlaGlyGlyLeuValAspLeuThrThrAlaThrAsnTh 942

5928 GAGAGATCCGAATTTG......GAAGGTCTCAAC...AGTGCTACCACAG 5968
                ::::|||            ||||||   ||||||||||||.
 942 rLeuGlyIleSerLeuAlaAspAlaAspLeuAsnValSerAlaThrThrV 959

5969 TTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTT 6018
     ||::::::   :::|||:::
 959 alAspThrAlaThrValSer.......................... 965

6019 TTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGT 6068
     |||   |||:::::::::            :::   |||:::|||   ::
 966 LeuLysAspSerAlaAsn............AsnSerLeuSerLeuThrLe 978

6069 CTTCGCAACCGGCGCGTCGGAAGGCCTCTTT 6099
     :      |||||||||:::   |||:::|||
 978 uValGluThrGlyAlaAsnThrGlyValPhe 988
```

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-362-525-22

```
seq_documentation_block:
; Sequence 22, Application US/08362525
; Patent No. 6027910
;  GENERAL INFORMATION:
;    APPLICANT:  KLIS, FRANCISCUS M.
;    APPLICANT:  SCHREUDER, MAARTEN P.
;    APPLICANT:  TOSCHKA, HOLSER Y.
;    APPLICANT:  VERRIPS, CORNELIS T.
;    TITLE OF INVENTION:  PROCESS FOR IMMOBILIZING ENZYMES TO THE
;    TITLE OF INVENTION:  CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
;    TITLE OF INVENTION:  PROTEIN
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,525
;      FILING DATE:  04-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92202080.5
;      FILING DATE:  08-JUL-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 92203899.7
;      FILING DATE:  14-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/EP93/01763
;      FILING DATE:  07-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KOKULIS, PAUL N.
;      REGISTRATION NUMBER:  16,773
;      REFERENCE/DOCKET NUMBER:  213289/T7020(V)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-362-525-22


alignment_scores:
            Quality:  149.50            Length:   1030
              Ratio:  0.340              Gaps:     54
 Percent Similarity:  42.718   Percent Identity: 20.388

alignment_block:
 US-09-301-906-3/rev x US-08-362-525-22   ..

 Align seq 1/1  to: US-08-362-525-22  from: 1  to: 894

    5504 GGTACTTTCAGAACACCTTCGGCGGTA.........GTGACGGCAATATT 5464
         |||   :::   ||||||:::   |||         :::|||:::   ||
     128 GlyPheTyrThrThrProThrAsnValThrLeuGluMetThrGlyTyrPh 144

    5463 TCTAGCCATAAGCAACGCCCTGTATGTCCTCTCAATGCCTATAGCCAGAT 5414
         ||||      :::::::::   |||:::
     144 cLeuProProGlnThrGlySerTyrThr..................... 153

    5413 ACTTCAAGTACTCCACGACGGCATCTTCAGAGGTC.............. 5379
           ||||||:::::::|||:::   :::|||   :::
     154 ..PheLysPheAlaThrValAspAspSerAlaIleLeuSerValGlyGly 169

    5378 ...............................CCGGTTTTAACCACTTC 5362
                                        |||   :::|||:::::
     170 AlaThrAlaPheAsnCysCysAlaGlnGlnGlnProProIleThrSerTh 186

    5361 AAATGTCATTGGGGTGTAGTCAGGTTCAAGAGCCATCACTGGATTATCAG 5312
         :|||                    :::|||   |||::::::
     186 rAsnPhe.....................ThrIleAspGlyIleLysP 195

    5311 CCCTGAGCCACTCA................................. 5298
              :::   |||
     195 roTrpGlyGlySerLeuProProAsnIleGluGlyThrValTyrMetTyr 211

    5297 ............CACTTCCTAACGTAATCTTGTTCGCACCATATCGAGTA 5260
                     :::   :::           |||:::   ::::::::
     212 AlaGlyTyrTyrTyrProMetLysValValTyrSerAsnAlaValSerTr 228

    5259 TGTCGACCTCCCCTTGACTTCCCTACCACTTCCCTCAGATAGAGCCTTGC 5210
         :     ||||||:::   |||:::   ||||||:::      ::::::
     228 pGlyThrLeuProIle...SerValThrLeuProAspGlyThrThrValS 244

    5209 CCCCCGATTTTGGTGTGATTGTTCTGGTAGCACCTTGGCTCTGTGGTTGC 5160
              ||||||
     244 erAspAspPheGlu.................................. 248

    5159 TGGTTTGGAACACCTGGTGACTTCTCATCCGACACGCCGCCGGCTAACTG 5110
               |||         :::|||::::::|||         :::|||||
     249 ......GlyTyrValTyrSerPheAspAspAspLeuSerGlnSerAsnCy 263

    5109 CAACATCCTGCCTACACCAAAGACGCGCCGAGTTGCCCTGAACGTCAAAC 5060
         |::::::   |||   |||::::::      |||:::   ::::::
     263 sThrVal...ProAspProSerAsnTyrAlaValSerThrThrThrThrT 279

    5059 TGATGCGCCCTTCGTCGATGGAGTTTACTGAATGCAAGTGCGTCCTTTGA 5010
              :::|||   :::       ||||||:::
     279 hrThrGluProTrpThrGlyThrPheThrSerThr.............. 290

    5009 AAACCGCAAGGCATAAGAAAATAGTCACCCGAAGCTACGTTTATGACCAT 4960
                                 |||   |||   |||   :::|||
     291 .......................SerThrGluMetThrThrValThrGl 299

    4959 GACC......TTACCACCCTTCCTGCACTTAGTCGAGAAGTTTGCCTTCC 4916
          |||      :::|||                              :::
     299 yThrAsnGlyValPro...........................ThrAspG 307

    4915 CCACGAGATTGACCGTCAAGATAGGGTTATCTGATGGATAGCACTCCTCG 4866
           |||   ::::::::::::         |||:::|||      |||:::
     307 luThrValIleValIleArgThrProThrSerGluGlyLeuIleSerThr 323

    4865 TCATCA...GCGTGGAATGGTACGCCTCCACCCATCTTGTACTTCTGCAC 4819
         ::::::      |||:::||||||                ::::      ||
     324 ThrThrGluProTrpThrGlyThr..............PheThrSerTh 335

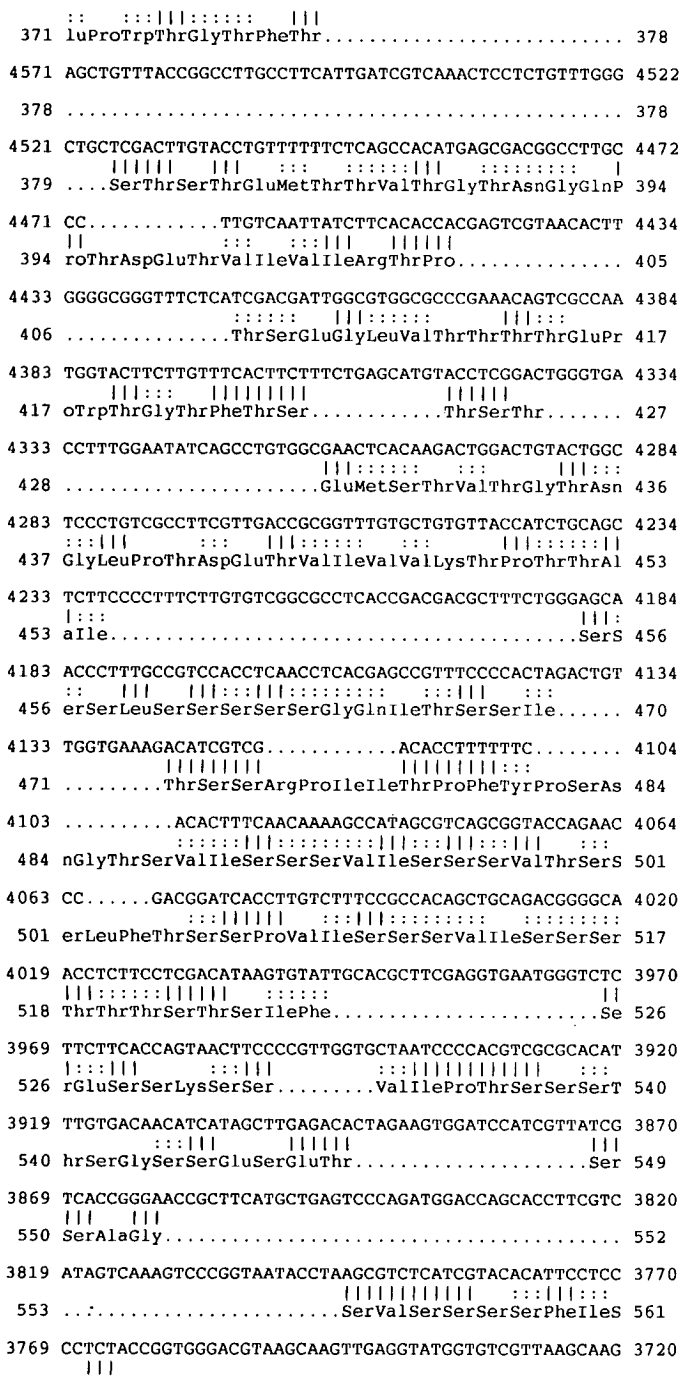
    4818 TAAACAGTGGTCGAAGACGCTTGGGTACTTAATTGCCGTTAAGATATCCT 4769
         |:::             |||:::                         :
     335 rSerThrGluValThrThrIle..........................T 343

    4768 CTAGGGCACGAGGCCACCCT...GATGAAACATGGCTACCACCATTATAG 4722
         ::   ::::::|||:::|||   |||||||||   :::
     343 hrGlyThrAsnGlyGlnProThrAspGluThrValIle............ 355

    4721 GTATACCCCTCACCCAGCTTTGAGAAAAAGACCGCACTCCTCCCTTTCAA 4672
         |||      :::|||:::
     356 ValIleArgThrProThr.............................. 361

    4671 GTTATCGACAAAGGTAAATGCGTTTGAGAACCTGGTGGCGATAGTCGCTT 4622
                                  |||:::|||::::::   :::::::
     362 .....................SerGluGlyLeuIleSerThrThrThrG 371

    4621 CGTTGTACACATCCAGCTGCACACCGCAGGAACACGTTCTAAACATGCAA 4572
```


luProTrpThrGlyThrPheThr
AGCTGTTTACCGGCCTTGCCTTCATTGATCGTCAAACTCCTCTGTTTGGG
CTGCTCGACTTGTACCTGTTTTTTCTCAGCCACATGAGCGACGGCCTTGC
....SerThrSerThrGluMetThrThrValThrGlyThrAsnGlyGlnP
CC.....TTGTCAATTATCTTCACACCACGAGTCGTAACACTT
roThrAspGluThrValIleValIleArgThrPro
GGGGCGGGTTTCTCATCGACGATTGGCGTGGCGCCCGAAACAGTCGCCAA
.....ThrSerGluGlyLeuValThrThrThrThrGluPr
TGGTACTTCTTGTTTCACTTCTTTCTGAGCATGTACCTCGGACTGGGTGA
oTrpThrGlyThrPheThrSer.....ThrSerThr.....
CCTTTGGAATATCAGCCTGTGGCGAACTCACAAGACTGGACTGTACTGGC
.....GluMetSerThrValThrGlyThrAsn
TCCCTGTCGCCTTCGTTGACCGCGGTTTGTGCTGTGTTACCATCTGCAGC
GlyLeuProThrAspGluThrValIleValValLysThrProThrThrAl
TCTTCCCCTTTCTTGTGTCGGCGCCTCACCGACGACGCTTTCTGGGAGCA
aIle.....SerS
ACCCTTTGCCGTCCACCTCAACCTCACGAGCCGTTTCCCCACTAGACTGT
erSerLeuSerSerSerSerSerGlyGlnIleThrSerSerIle.....
TGGTGAAAGACATCGTCG.....ACACCTTTTTTC.....
.....ThrSerSerArgProIleIleThrProPheTyrProSerAs
.....ACACTTTCAACAAAAGCCATAGCGTCAGCGGTACCAGAAC
nGlyThrSerValIleSerSerSerValIleSerSerSerValThrSerS
CC.....GACGGATCACCTTGTCTTTCCGCCACAGCTGCAGACGGGGCA
erLeuPheThrSerSerProValIleSerSerSerValIleSerSerSer
ACCTCTTCCTCGACATAAGTGTATTGCACGCTTCGAGGTGAATGGGTCTC
ThrThrThrSerThrSerIlePhe.....Se
TTCTTCACCAGTAACTTCCCCGTTGGTGCTAATCCCCACGTCGCGCACAT
rGluSerSerLysSerSer.....ValIleProThrSerSerSerT
TTGTGACAACATCATAGCTTGAGACACTAGAAGTGGATCCATCGTTATCG
hrSerGlySerSerGluSerGluThr.....Ser
TCACCGGGAACCGCTTCATGCTGAGTCCCAGATGGACCAGCACCTTCGTC
SerAlaGly.....
ATAGTCAAAGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCC
..:.....SerValSerSerSerSerPheIleS
CCTCTACCGGTGGGACGTAAGCAAGTTGAGGTATGGTGTCGTTAAGCAAG
erSer.....
CCCCTTATGATCAAAGATACCCCTAGCGTAATAGGCAGTAACAATGAGCT
.....GluSerSerLysSerProThrTyrSerSerSerSerLe
GACACTCTTCCCTCGCCAAACAACCTCGGGAACAGCGCACCAATTATTGG
uProLeuValThrSerAlaThrThrSer.....
AGAAGCCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCAAGCAAT
.....GlnGluThrAlaSerSerLeuProProAlaThr
GTCCCTAAACCAAAGAAAAAC.....ATTAAGTGCCG
ThrThrLysThrSerGluGlnThrThrLeuValThrValThrSerCysGl
TGGAAGAGCATGTGCCCAAATGCCTAACATACTGCAGCCGGAAATGGCGT
uSerHisValCysThrGluSer.....IleSerProAlaIleVal.
ATGAGGTAGAAAACGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACG
.....SerThrAlaThr
GTCGCCCTAGTGACGGTCACCGCCGACGAAAAC.....CCTCTTCG
ValThrValSerGlyValThrThrGluTyrThrThrTrpCysProIleSe
CACGACTAAAGATCCACCGGAT.....ACTGTGCTAGAGGTTA
rThrThrGluThrThrLysGlnThrLysGlyThrThrGluGlnThrThrG
AACCGCGTAAACCACCGGCGGCTTGTGAACTTAACGAGCTTAGAGTTTCG
luThrThrLys.GlnThrThrValValThrIleSerSerCysGlu...Se
TAACTACTGCCCGCGTACCTTCTAGC.....GTTAGCCGCAATTCC
rAspValCysSerLysThr.AlaSerProAlaIleValSerThrSerThr
GCGCCTATACTTCCATGCAATTGGGGCCGCACCCAAAGACGCCACGAATA
AlaThrIle.....Asn.GlyValThrThrGluTyrThrThrTrpC
TCCCGATAAGCACCTTATTCCGCAATACGAAAGTGGTGAAGTCATTTAAG
ysProIleSerThr.....Thr
GAAACTTCACTCCAAAATTCTTTCGTTCCAGCTACGACCTTGCTCACAGC
GluSerArgGlnGln.....ThrThrLeuValThrVa
TACGCGCGACGTCTCCAGTATGCCCACCGGTGAGTTTATAGCATTCTCTG
lThrSerCysGluSerGlyValCysSerGluThrAlaSerProAlaIleV
TGCGCACCAGAGCATCAAGCCTTTCATCTTCAGTT.....
alSerThrAlaThrAlaThrValAsnAspValValThrValTyrProThr
...AAACCGCGGTCGTAGAATAGTGAACCTGTG.....GAAAAGAACGA
TrpArgProGlnThrAlaAsnGluGluSerValSerSerLysMetAsnSe
CGTAATGCCAGAAAAGGCC.....TTCGTCGCAGCCCCACGAGCTACCA
rAlaThrGlyGluThrThrThrAsnThrLeuAlaAlaGluThrThrThrA
AACCCTTA.....GAAACACTCGAAACCACGGTCGCTTTCGAACCGCCA
snThrValAlaAlaGluThrIleThrAsnThrGlyAlaAlaGlu.....

```
2903 AGAAGACCTGCCCTCGTACAGTGGACGTTAGAAACGTTCCCCATTATGAA 2854
           ::::::|||      :::|||:::               ::
 806 ...ThrLysThrValValThrSerSerLeuSerArgSerAsnHisAlaGl 821

2853 CGCATATACAGAATCTTCA......GTAATTGGCAGGGGTTTTCCATCTT 2810
     ::::   |||   |||:::      |||||||||:::
 821 uThrGlnThrAlaSerAlaThrAspValIleGlyHis..... 833

2809 CTGTCTTCATTGCCTTCATGACCTTAATTGAATCTTCCACGATAGCCCTA 2760
                                :::||||||:::::::::::::
 834 ..........................SerSerSerValValSerVal 840

2759 GAGAAGACG......CGCAAGCGACTATTCTTCAGTTTCCTACCGAAGGT 2716
     :::::::|||         |||   |||      |||   |||      ::
 841 SerGluThrGlyAsnThrLysSerLeuThrSerSerGlyLeuSerThrMe 857

2715 CTCCTTCGAAATATCAAAGGCACCCCTGAGCTCC 2682
     :|||   :::      ::::::|||   ||||||
 857 tSerGlnGlnProArgSerThrProAlaSerSer 868
```

```
OM of: US-09-301-906-3 to: PIR_66:*   out_format : pfs

Date: Jan 31, 2001  3:49 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09301906/runat_31012001_153035_9612/app_query.fasta_1.6881
 -DB=PIR_66 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09301906_@CGN1_1_784 -NCPU=6 -ICPU=3
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-3
Query length: 6714
Database: PIR_66:*
Database sequences: 195891
Database length: 67900655
Search time (sec): 527.440000


score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation  ..
pir2:A49282         +  629.00  718.37  2.6e-33   1729  ! fusion protein 1a/1b - sugar b
pir1:JQ1734         +  414.00  462.66  4.6e-19   1718  ! genome polyprotein - shallot v
pir2:JQ0096         +  404.00  451.97  2.0e-18   1547  ! hypothetical 176K protein - pa
pir2:S77910         +  394.00  438.94  9.6e-18   1707  ! hypothetical protein 1 - grape
pir2:S77908         +  391.50  435.96  1.4e-17   1707  ! hypothetical protein 1 - grape
pir2:S01865         +  338.50  372.91  4.6e-14   1707  ! genome polyprotein - tobacco r
pir1:WMFM12         +  329.50  366.98  1.5e-13   1126  ! 125K protein - alfalfa mosaic
pir2:T45025         -  318.00  340.04  1.5e-12   3570  ! mucin MUC5B, tracheobronchial
pir1:WMTM8T         +  309.00  338.45  4.0e-12   1615  ! 180K protein - tomato mosaic v
pir2:S03701         +  306.00  333.85  6.7e-12   1766  ! 141K protein - pea early brown
pir1:WMTMPV         +  297.00  324.20  2.5e-11   1611  ! 183K protein - pepper mild mot
pir1:WMTMS2         +  294.50  320.98  3.7e-11   1646  ! 186K protein - cucumber green
pir2:JQ1896         +  291.00  316.49  6.4e-11   1694  ! hypothetical 191.1K protein -
pir1:P1BVBB         +  290.00  321.75  5.7e-11    966  ! RNA 1a protein - broad bean mot
pir2:JQ2183         +  281.00  303.36  3.1e-10   1885  ! hypothetical 216.5K protein -
pir2:JA0074         +  277.00  305.97  4.2e-10    993  ! RNA 1 protein - cucumber mosaic
pir1:P1VXPJ         +  275.50  304.05  5.3e-10   1005  ! RNA 1 protein - peanut stunt v
pir1:P1BVA          +  275.00  303.97  5.6e-10    961  ! RNA 1a protein - brome mosaic v
pir1:PAVBBS         +  273.50  300.23  7.7e-10   1139  ! alpha-a protein - barley strip
pir1:P1VXCM         +  273.00  301.21  7.8e-10    993  ! RNA 1 protein - cucumber mosaic
pir1:WMTMGM         +  272.50  295.08  1.1e-09   1608  ! 183K protein - tobacco mild gr
pir2:S25345         -  270.50  292.69  1.4e-09   1609  ! probable membrane protein YCR0
pir2:E71392         +  269.00  296.45  1.4e-09    993  ! ATP-dependent RNA helicase - cu
pir1:WMTM18         +  267.00  288.48  2.4e-09   1615  ! 183.3K protein - tobacco mosai
pir1:JQ2144         +  264.50  285.50  3.6e-09   1616  ! 183K protein - tomato mosaic v
pir2:S26358         +  264.00  289.16  3.2e-09   1116  ! hypothetical protein, 126K - t
pir1:P1BVCV         +  263.00  289.34  3.6e-09    991  ! RNA 1 protein - cucumber mosaic
pir1:A45353         +  261.00  279.57  6.6e-09   1884  ! genome polyprotein - apple chl
pir2:T34434         -  261.00  277.63  7.1e-09   2232  ! hypothetical protein K06A9.1a
pir2:JQ2169         +  253.50  278.01  1.5e-08    993  ! RNA 1 protein - cucumber mosaic
pir1:P1BVCC         +  252.00  276.64  1.9e-08    958  ! 1a protein - cowpea chlorotic m
pir2:S49432         +  252.00  274.92  2.0e-08   1112  ! replicase 126K - odontoglossum
pir2:S57180         -  241.50  261.94  1.0e-07   1161  ! probable membrane protein YJR1
pir1:S48478         -  240.00  258.28  1.4e-07   1367  ! glucan 1,4-alpha-glucosidase (
pir2:T30826         -  239.50  252.28  1.9e-07   2187  ! nascent polypeptide-associated
pir1:P1VXTA         +  237.00  258.38  1.9e-07    993  ! RNA 1 protein - tomato aspermy
pir2:S48699         +  236.00  251.70  2.8e-07   1601  ! 178K protein - tobacco mosaic
pir2:S59310         -  224.50  242.29  1.3e-06   1104  ! probable membrane protein YMR3
pir2:T34513         -  219.00  222.47  5.4e-06   3507  ! hypothetical protein ZK783.1 -
pir2:S65053         +  218.00  230.32  4.3e-06   1597  ! genome polyprotein - Chinese r
pir2:A53577         -  209.50  219.97  1.6e-05   1630  ! ascites sialoglycoprotein 1 -
pir2:T34369         -  206.00  214.81  2.8e-05   1777  ! hypothetical protein T19D12
pir2:T42233         +  203.00  212.53  4.2e-05   1589  ! submaxillary mucin 1 - bovi
pir2:A41258         -  198.50  216.19  5.8e-05    725  ! a-agglutinin core protein AG
pir2:S25370         -  195.50  205.86   0.0001   1306  ! MSB2 protein - yeast (Sacch

seq_name: pir2:A49282

seq_documentation_block:
fusion protein 1a/1b - sugar beet yellows virus (isolate BYV-U) (fragments)
C;Species: sugar beet yellows virus, SBYV
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 17-Mar-1999
C;Accession: A49282
R;Agranovsky, A.A.; Koonin, E.V.; Boyko, V.P.; Maiss, E.; Froetschl, R.; Lunina, N.A.
Virology 198, 311-324, 1994
A;Title: Beet yellows closterovirus: complete genome structure and identification of
A;Reference number: A49282; MUID:94082464
A;Accession: A49282
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-1729 <AGR>
A;Cross-references: GB:X73476
A;Note: this protein is predicted to result from +1 ribosomal frameshifting during tr


alignment_scores:
           Quality:  629.00              Length:   1862
             Ratio:  0.834                Gaps:     76
 Percent Similarity:  40.494   Percent Identity: 20.838

alignment_block:
 US-09-301-906-3 x A49282   ..

 Align seg 1/1  to: A49282  from: 1  to: 1729

   1285 TTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGGGA 1334
           |||         ::::::      |||   |||:::   ::::::
      4 LeuAlaArgSerValGluLysAspLeuIleAspPheLysAspGluIleLy 20

   1335 CGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTTCTGAAGC 1384
         ::::::   :::||||||   |||                 ::::::
     20 sSerLeuSerLysGluLysArgSerValThrValProPheTyrMetGlyG 37

   1385 CGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAG 1434
                ::::::   |||   :::   :::|||:::::::::::::::
     37 luAlaValGlnSerGlyLeuThrArgAlaTyrProGlnPheAsnLeuSer 53

   1435 TTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAG 1484
           |||:::   ||||||   |||:::||||||   |||   :::   |||
     54 PheThrHisSerValTyrSerAspHisProAlaAlaAlaGlySerArgLe 70

   1485 CTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCG 1534
              |||      ::::::      :::   :::      |||   |
     70 uLeuGluAsnGluThrLeuAlaSerMetAlaLysSer...SerPheSerA 86

   1535 ATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGT 1584
           ||||||||||          :::|||:::|||   |||   ::::::
     86 spIleGlyGlyCysProLeuPheHisIleLysArgGlySerThrAspTyr 102

   1585 CATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATCAA 1634
           |||||||||:::|||:::   |||:::||||||::::::||||||:::::
    103 HisValCysArgProIleTyrAspMetLysAspAlaGlnArgArgValSe 119

   1635 TGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGTG 1684
           ::::         |||:::::   |||       :::   :::|||   :
    119 rArg......GluLeuGlnAlaArgGlyLeuValGluAsnLeuSerArgG 134

   1685 ACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAA 1734
           ::   |||:::||||||   ::::::   ||||||   |||   :::
    134 luGlnLeuValGluAlaGlnAlaArg...ValSerValCysProHisThr 149

   1735 TCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTTATGGTGGATGTGTA 1784
              |||||||:::   ::::::|||:::   :::||||||:::|||||
    150 LeuGlyAsnCysAsnValLysSerAspValLeuIleMetValGlnValTy 166
```

```
1785 CGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGC 1834
     ||||   |||   ::::::::::|||:::|||||      |||   :::
 166 rAspAlaSerLeuAsnGluIleAlaSerAlaMetValLeuLysGluSerL 183

1835 TGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTACGGTAAC 1884
       |||      ::::::::::::   |||   |||||||||      :::
 183 ysValAlaTyrLeuThrMetValThrProGlyGluLeuLeuAspGluArg 199

1885 GGTGAAGTTTACTTGGAAGAACTC......GATACGTTGGTGAAGAGGGA 1928
                  ::::::   |||      |||::::::|||      ::
 200 GluAlaPheAlaIleAspAlaLeuGlyCysAspValValValAspThrAr 216

1929 AGGTGATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAAC 1978
     :   |||   :::   |||:::   |||      ::::::   |||   |
 216 gArgAspMetValGlnTyrLysPheGly......SerSerCysTyrCysH 231

1979 ATTCCTTCTCTAACGTAAGCGGGTTT..............TTCACCTTT 2013
     ||::::::||||||::::::::::::             |||||||||
 231 isLysLeuSerAsnIleLysSerIleMetLeuThrProAlaPheThrPhe 247

2014 TCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGG 2063
                     |||||||||:::|||::::::|||
 248 ..................SerGlyAsnLeuPheSerValGluMetTyrGl 258

2064 ATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCAC 2113
       |||   |||   ::::::      :::   |||:::      ||||
 258 uAsnArgMetGlyValAsnTyrTyrLysIleThrArgSerAlaTyrSerP 275

2114 CTGGA.....................................ACTGAG 2124
     ||                                       ||||||
 275 roGluIleArgGlyValLysThrLeuArgTyrArgArgAlaCysThrGlu 291

2125 GTTACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTT 2174
     |||:::      :::|||   |||   |||                :::||
 292 ValValGlnValLysLeu...ProArgPheAspLysThrLeuLysThrPh 307

2175 CATACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCG 2224
     |:::                          ::::::      |||   ||||
 307 eLeuSer.....................GlyTyrAspTyrIleTyrLeuA 317

2225 ATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGG 2274
     ||:::   ||||||:::|||:::::::::|||:::::::::::::   :::
 317 spAlaLysPheValSerArgValPheAspTyrValValSerAsnCysSer 333

2275 ACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCA 2324
     :::   :::::::::|||||||||::::::   :::   ::::::|||::
 334 ValValAsnSerLysThrPheGluTrpValTrpSerTyrIleLysSerSe 350

2325 AAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATA 2374
     :|||::::::|||:::   :::::::::|||:::|||      |||   |
 350 rLysSerArgValValIleSerGlyLysValIleHisArgAspValHisI 367

2375 TTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCG 2424
     ||:::                :::::::::::::::::|||   :::
 367 leAspLeuLysHisSerGluCysPheAlaAlaValMetLeuAlaValGly 383

2425 ATTAAGGATAGG.................GCGAAGAGTATTCGCTCCTA 2456
     :::::::::|||                |||||:::::::::   ||
 384 ValArgSerArgThrThrThrGluPheLeuAlaLysAsnLeuAsnTyrTy 400

2457 TAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCT 2506
     |
 400 r..................................................  400

2507 TCTTTCAGACCGTAGGCGAT......TGTTTTTCG......AACGCAGTC 2544
                  :::||||||      ||||||:::      :::|||:::
 401 ...........ThrGlyAspAlaSerCysPheAlaLeuArgArgAlaLeu 413

2545 TCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGAC 2594
     |||:::|||:::
 414 SerAlaTyrSer..................................... 417
```

```
2595 GCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTG 2644

 417 ..................................................  417

2645 TTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGT 2694

 417 ..................................................  417

2695 GCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCG 2744
                                                     ||||||
 418 ..................................................AsnSerAl 420

2745 CTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGA 2794
      ||||||          |||   |||:::|||
 420 aLeuArg.........ArgLysIleIleGlu..................... 427

2795 AGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCT 2844

 427 ..................................................  427

2845 GTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGC 2894
             ||||||:::||||||:::   :::
 428 .........PheIlePheGlyAsnIleHisHisPro.............. 436

2895 AGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGG 2944

 436 ..................................................  436

2945 GTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCG 2994

 436 ..................................................  436

2995 TTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGA 3044
                              :::|||
 437 .........................PheAsp..................... 438

3045 AAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGG 3094
                                                       |||:
 439 ..................................................ValA 440

3095 GCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACG 3144
     ::::::::|||||:::::|||||
 440 laValIleGluThrAsnGluValAla.....................Pro 449

3145 AAAGAATTTTGGAGTGAAGTTTCCTTAAAT.........GACTTCACCAC 3185
     :::   :::       ||||||:::::::::        ||||||
 450 GluProLeuSerProGluValAspIleAspValAspCysAspPhe..... 464

3186 TTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGG 3235

 464 ..................................................  464

3236 GTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCT 3285

 464 ..................................................  464

3286 AGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCG......TTAAG 3329
                  ||||||:::   |||::::::||||||      :::::
 465 ............GlySerAspSerGluSerValSerSerAspGluValAl 477

3330 TTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCG 3379
     :|||:::         ||||||:::                        |
 477 aSerAsnProArgProGlyLeuHis.......................G 486

3380 GTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACT 3429
     ||||||||            |||      |||:::   :::|||
 486 lyGlySer............ArgArgSerSerAsnPheLeuThrSerLeu 498

3430 AGGGCGACCGTAGCTAAA......CGTCAAGTCCCCTTAGCGTTGCTATC 3473
           :::|||   |||      |||:::::::|||      |||:::
 499 ValLysValValPheLysLeuAlaArgArgIleProArgLeuLeuPheAr 515

3474 GTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGG 3523
```

```
      :::   :::      |||                        :::
  515 gLeuArgAsnPheValAlaTyrPhe.................ValGluA 526

 3524 CACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTG 3573
        :::         ::::::|||      |||:::|||||||:::   |||
  526 rgArgLeuAlaSerLysArgLeuLysThrPheIleGlyLeuAlaArgLeu 542

 3574 CTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAA 3623
                                      |||   :::||||||
  543 ................................PheAspAsnPheSer.. 547

 3624 TAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCT 3673
                   :::   :::||||||:::      ::::::   :::|
  548 .............LeuThrSerValValTyrLeuLeuGlnGluTyrAspS 560

 3674 CATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTG 3723
      ||:::|||      :::   :::   |||:::||||||         |||
  560 erValLeuAsnAlaPhe...IleAspValGluLeuIle.........Leu 572

 3724 CTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAG 3773
      ||||||
  573 LeuAsn..................................................... 574

 3774 GAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACG 3823

  574 ..................................................... 574

 3824 AAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGAT 3873

  574 ..................................................... 574

 3874 AACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGT 3923
                 ::::::||||||:::       :::|||:::   ||
  575 ............SerGlySerValAsnValLeuProLeuValSerTrpVa 587

 3924 GCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGA 3973
      ||||                  |||:::::|||         |||
  587 lArg.....................GlySerLeuThrLysLeuAlaGlu. 596

 3974 CCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCC 4023

  596 ..................................................... 596

 4024 CCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTAC 4073
             |||::::::                  |||||||||
  597 .........AlaIleVal.....................GlySerGlyPh 603

 4074 CGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATG 4123
       |||         :::|||:::      :::         |||:::|||
  603 eAla.........SerPheLeuGlyArgMetCysCysArgValSerAspT 617

 4124 TCTTTCACCAACAGTCTAGTGGG......GAAACGGCTCGTGAGGTTGAG 4167
             ::::::|||:::::::      :::             |||:::
  617 rpCysSerSerSerSerAsnAlaGlyCysAsnPheMetSerProValArg 633

 4168 GTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGAC 4217
      :::   ||||||   :::   |||:::|||      :::   |||
  634 ThrLysGlyLysPheValProProSerSerSerGlySerThrAlaSerMe 650

 4218 ACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCA 4267
          ||||||   :::|||   :::::::::
  650 tTyrGluArgLeuGluAlaLeuGluSerAspIle................ 661

 4268 ACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAG 4317
                ||||||   |||   |||:::
  662 ...........ArgGluHisValLeuSerThrCys.............. 669

 4318 GCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGT 4367
                   :::|||   :::::::|||
  670 ............ArgValGlySerAspGlu..................... 675

 4368 GAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCG 4417
```

```
  675 ..................................................... 675

 4418 ATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGAC 4467

  675 ..................................................... 675

 4468 AAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGA 4517
                                                  :::   ||
  676 .....................................GluGluGl 678

 4518 GCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAAC 4567
      |:::||||||:::
  678 uArgProLysGlu.................................... 682

 4568 AGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTAC 4617

  682 ..................................................... 682

 4618 AACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGA 4667

  682 ..................................................... 682

 4668 TAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGT 4717

  683 .....................................................V 683

 4718 ATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTA 4767
        |||       |||   :::|||:::|||
  683 alThrGluProGlyIleGluHisThrSer..................... 692

 4768 GAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTT 4817
      ||||||::::::                        :::   :::|||
  693 GluAspValVal.................ProIleArgSerHis..... 701

 4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACG 4867
             :::       ::::::|||||         ::::::::::|||:
  702 .......SerGlnProLeuSerGlyGlyGluCysSerTyrSerGluAspA 716

 4868 AGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGG 4917
      ::|||             :::|||
  716 rgGlu............GluAsnGlu....................... 720

 4918 AAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCAT 4967
      :::||||||
  721 ArgAlaAsn.......................................... 723

 4968 AAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGA 5017
                                |||:::|||
  724 .........................LeuLeuPro................. 726

 5018 CGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACG 5067
        |||:::   ::::::   |||      :::|||   ::::::   :::
  727 ..HisValSerLysIleValSerGluArgArgGly...LeuGluThrAla 741

 5068 TTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGC 5117
        |||       |||   :::   ||||||::::::::::|||:::
  742 ArgArgAsnLysArgThrLeuHisGlyValSerGluPheLeuAsnAlaIl 758

 5118 CGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCAC 5167
       :::       :::::::::       |||   :::   ::::::::::|||:
  758 eAsnThrSerAsnGluGlnProArgProIleIleValAspHisSerProG 775

 5168 AGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCT 5217
      ::|||::::::   |||::::::::::
  775 luSerArgAlaLeuThrAsnSerVal....................... 783

 5218 CTATCTGAGGGAAGTGGTAGGGAA.........GTCAAGGGGAGGTCGAC 5258
                          ||||||       ::::::      :::
  784 ..................ArgGluPheTyrTyrLeuGlnGluLeuAlaLe 794

 5259 ATACTCGATATGGTGCGAA...CAAGATTACGTTAGGAAGTGTGAGTGGC 5305
       :::::::::   |||:::   ::::::|||      :::   :::
  794 uPheGluLeuSerCysLysLeuArgGluTyrTyrAspGlnLeuLysValA 811
```

```
5306 TCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATG 5355
811 laAsnPheAsnArgGlnGluCysLeuCysAspLysAsp..... 823

5356 ACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTA 5405
824 .....GluAspMetPheVal..... 828

5406 CTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTA 5455

828 ..... 828

5456 TGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCT 5505
829 .....LeuArgAlaGlyGlnGlyValValSerGlyArg 839

5506 AATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCAC 5555
840 Asn.....SerArgLeuProLeuLysHisPhe.....LysGlyHi 851

5556 AGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTAC 5605
851 sGluPheCysPheArgSer.....GlyGlyLeu.....ValP 862

5606 CGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTC 5655
862 roTyr..... 863

5656 GACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACT 5705
864 .....AspGlyThrSerArgValAspThrIle.. 872

5706 GGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGAT 5755
873 ....PheHisThrGlnThrAsnPheValSerAla...AsnAlaLeuLeuS 887

5756 GCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAG..... 5796
887 erGlyTyrLeu.....SerTyrArgThrPheThrPheThrAsnLeuSer 901

5797 .....TGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGAC 5834
902 AlaAsnValLeuLeuTyrGluAlaProPro...GlyGlyLysThrThrTh 917

5835 GTTAGTGGACGAATTCGTTAAG.....TCACCCAATAGCACAGCCACCA 5878
917 rLeuIleLysValPheCysGluThrPheSerLysValAsnSerLeuIleL 934

5879 TTACGGCTAATGTGGGAAGTTCTGAGGACATA.....AAT 5913
934 euThrAlaAsnLysSerSerArgGluGluIleLeuAlaLysValAsnArg 950

5914 ATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTAC 5963
951 IleValLeuAspGluGlyAspThrProLeuGlnThrArgAspArgIleLe 967

5964 CACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAA 6013
967 uThrIleAspSerTyrLeuMetAsn.....AsnArgGlyLeuThrCysL 982

6014 GGGTTTTG...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAA 6060
982 ysValLeuTyrLeuAspGluCysPheMetValHisAlaGlyAlaAlaVal 998

6061 CTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACAT 6110
999 AlaCysIleGluPheThrLysCysAspSerAlaIleLeuPheGlyAspSe 1015

6111 AAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTG 6160
1015 rArgGlnIleArgTyr.....GlyArgCysSerGluLeuAspThrA 1029

6161 CTGTT.....TTTGTTCCAAAGAAGGAAAGCGTTGTA 6192
1029 laValLeuSerAspLeuAsnArgPheVal...AspAspGluSerArgVal 1044

6193 TACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTC 6242
1045 TyrGlyGluValSerTyrArgCysProTrpAspValCysAlaTrpLeuSe 1061

6243 CTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTT. 6291
1061 rThrPhe.....TyrProLysThrValAlaT 1070

6292 .....AGCGGTAAGGACAAACCAGTAGTAAGATCGCTG 6324
1070 hrThrAsnLeuValSerAlaGlyGlnSer.....SerMet 1081

6325 TCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGA 6374
1082 GlnValArgGluIleGluSerValAspAspValGluTyrSerSerGluPn 1098

6375 CGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATG..... 6414
1098 eValTyrLeuThrMetLeuGlnSerGluLysLysAspLeuLeuLysSerP 1115

6415 .....AAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACA 6459
1115 heGlyLysArgSerArgSerSerValGluLysProThr...ValLeuThr 1130

6460 GTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGAC 6509
1131 ValHisGluAlaGlnGlyGluThrTyrArgLysValAsnLeuValArgTh 1147

6510 GAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTG 6559
1147 rLysPheGlnGluAspAspProPheArgSerGluAsnHisIleThrValA 1164

6560 GTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAG 6609
1164 laLeuSerArgHisValGluSerLeuThrTyrSerValLeuSerSerLys 1180

6610 TTGGACGATAAGGTCGGCACATATATTAGCGACGCG 6645
1181 ArgAspAspAlaIleAlaGlnAlaIleValLysAla 1192
```

seq_name: pir1:JQ1734

seq_documentation_block:
genome polyprotein - shallot virus X
N;Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: shallot virus X
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 11-Jun-1999
C;Accession: JQ1734
R;Kanyuka, K.V.; Vishnichenko, V.K.; Levay, K.E.; Kondrikov, D.Y.; Ryabov, E.V.; Zavr
J. Gen. Virol. 73, 2553-2560, 1992
A;Title: Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal cistron clo
A;Reference number: JQ1734; MUID:93019008
A;Accession: JQ1734
A;Molecule type: genomic RNA
A;Residues: 1-1718 <KAN>
A;Cross-references: GB:M97264; NID:g295078; PIDN:AAA47787.1; PID:g295079
C;Superfamily: eggplant mosaic virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase; P-loop; RNA biosynthesis; RNA replication
F;915-922/Region: nucleotide-binding motif A (P-loop)
F;978-983/Region: nucleotide-binding motif B
F;921/Binding site: ATP (Lys) #status predicted

alignment_scores:
Quality: 414.00 Length: 1025
Ratio: 0.828 Gaps: 51
Percent Similarity: 48.780 Percent Identity: 22.244

```
alignment_block:
 US-09-301-906-3 x JQ1734   ..

 Align seg 1/1  to: JQ1734  from: 1  to: 1718

    3974 CCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCC 4023
         |||:::|||:::            |||   :::   :::::::|||
     411 ProValHisPhe............ThrThrGlnProLysArgArgGluAl 423

    4024 CCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTAC 4073
                ::::||||||   ::::::
     423 aValGlyGluLeuTrpPheGlnGlu..................... 431

    4074 CGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATG 4123
                     |||                    :::||||||||||||
     432 .....ProLysTrp.....................SerValSerThrMet 439

    4124 TCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGAC 4173
              |||:::                 :::   :::::::::   |||||
     440 ......ThrGlnProArgLysGluHisHisArgLeuGlnMetThrTrpTh 454

    4174 G.......................................GCAAAGG 4181
         |                                        ::::::
     454 rLeuLeuAlaTrpPheHisGlnLeuGluSerSerGlySerMetSerGluP 471

    4182 GTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGC.........CGACACAAG 4222
           |||::::::::::   |||:::   :::|||          ::::::
     471 roCysAsnAsnSerGluSerThrProGlnArgThrAlaThrSerGlnGln 487

    4223 AAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAA 4272
         |||:::   :::|||            :::|||:::   |||:::   ::
     488 LysAlaAlaLysLeuThrThrSerGlnLysHisAsnArgArgThrAspGl 504

    4273 GGCGACA...GGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGC 4319
         ::::|||   :::   ||||||   |||                    |
     504 nThrThrMetAsnProGlnTyrProPro.....................L 514

    4320 TGATATTCCAAAGGTCACCC...AGTCCGAGGTACATGCTCAGAAAGAAG 4366
         ||:::::::      :::|||      |||||||::   |||   ||||||
     514 euMetLeuThrIleAlaProMetMetProArgHisSerLeuMetLysLys 530

    4367 TGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATC... 4413
                  :::|||      |||:::::::      :::   :::
     531 ..ThrIleAlaThrProCysArgThrLeuGluGluIleSerAspLeuAsp 546

    4414 .........GTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGT 4454
                     |||:::   |||      :::   :::
     547 LeuAspAspPheAspAspLeuProAsnGluAlaSerAsn........... 559

    4455 GAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAAC 4504
                                                 |||      :
     560 .....................................GluProProS 563

    4505 AGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGC 4554
         ::::::::   :::::::|||   :::::::::   |||   :::::::|||
     563 erAlaAsnGluGlnSerProAspAsnHisAlaGluThrThrThrArgGly 579

    4555 AAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCA 4604
                           :::|||      |||:::||||||:::::
     580 ....................ValPhe...ProCysGluCysGlyThrGl 588

    4605 GCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCAT 4654
         ::::   |||                     :::   |||:::::::|||:
     588 uIleThrVal.....................AsnSerPheGlyArgAlaI 598

    4655 TT..............ACCTTTGTCGATAACTTGAAAGGGAGGAGTGCG 4689
         ::              :::::::::|||:::::::|||||||||   |||
     598 leGluValAlaGlyValAsnLeuThrAspHisMetLysGlyArgLeuAla 614

    4690 GTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCA 4739
         :::|||:::|||:::   |||:::||||||:::|||:::|||   |||||
     615 AlaPheTyrSerArgAspGlyGlnGlyTyrSerTyrThrGlyTyrSerHi 631

    4740 TGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATT. 4788
         |   |||:::||||||   ::::::|||:::   ::::::   |||
     631 sLysSerGlnGlyTrpLeuGluGlyLeuAspLysLeuIleGluAlaCysG 648

    4789 ..AAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATG 4836
           :::   |||:::::::::::::::::|||||||||||||||||||:::::
     648 lyGluLysProThrThrTyrAsnGlnCysLeuValGlnLysTyrGluGln 664

    4837 GGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGA 4886
         |||:::   :::   ||||||:::|||::::::      |||||||:::
     665 GlySerArgIleGlyPheHisSerAspGluGlnAlaIleTyrProLysGl 681

    4887 TAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTA 4936
          |||   ||||||||||||||   :::|||::::::::|||:::   |
     681 yAsnLysIleLeuThrValAsnAlaAlaGlySerGlyThrPheGlyIleL 698

    4937 AGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGAC 4986
         |||||   ||||||            ::::::|||:::   :::||||||
     698 ysCysAlaLysGly......GluThrThrLeuAsnLeuGluAspGlyAsp 712

    4987 TATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAA 5036
         ||||||   ||||||   |||||||||:::||||||   |||:::|||
     713 TyrPheGlnMetProSerGlyPheGlnGluThrHisLysHisAsnValVa 729

    5037 CTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCG 5086
         ::::::          |||:::|||:::|||||||||:::|||
     729 lAlaValThrPro...ArgLeuSerPheThrPheArgSerThr....... 742

    5087 TCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAG 5136
                                                |||:::::::::
     743 .................................ValValAsnSerGln 747

    5137 AAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAG 5186
         |||:::|||:::   |||::::::      |||:::::::::::::   ::
     748 LysLysProAlaGluProGluLysLeuAsnGlnAsnAsnAlaCysProLy 764

    5187 AACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTA 5236
         :       |||                        |||:::::::||||||:
     764 sProSerAspPro.....................SerAsnAlaSerGlyL 774

    5237 GGGAA..................GTCAAGGGGAGGTCGACATACTCGATA 5268
         :::::                    :::|||||||::::::
     774 ysGlnHisLysLysThrHisProAlaLysGlyAsnGlu............. 786

    5269 TGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAA 5318
                                                 ::::::::::
     787 ..................................LysSerSerSe 790

    5319 TCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGG 5368
         :|||         |||||||||      :::|||   |||   |||::::
     790 rPro......AsnLeuGluProLeuAspAlaPro...ThrValGluIleL 804

    5369 TTAAA................................ACCGGGACCTCT 5385
         ::|||                                  ||||||
     804 euLysLeuHisGlyPheThrAlaLeuThrProGlnHisAspGlyThrCys 820

    5386 GAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAG 5435
         :::          |||   :::   |||          |||   ::::::||
     821 GlnIleArgProValTyrPheAsnLys......AspIleHisLeuArgAr 835

    5436 GACATACAGGGCGTTGCTTATG.........GCTAGA.........AATA 5467
         |                    |||          ||||||          ::::
     835 gLysAlaValLysThrAspMetSerProProAlaArgProPhePheAspL 852

    5468 TTGCCGTCACTACCGCCGAAGGTGTTCTG......AAAGTACCTAATCAA 5511
         ::|||::::::       :::|||:::          |||:::   |||:::
     852 euAlaThrSerLeuHisArgGlyIleTyrThrHisLysIleAspAsnArg 868

    5512 .......................GTTTATGAATCACTACCGGGCTTTCA 5537
                              |||   ::::::|||   |||:::
     869 ArgAlaThrAlaTyrMetSerAspValLysAsnAsnLeuThrGlyLeuVa 885

    5538 CGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCT 5587
```

```
      :::   |||         ||||||:::      |||:::      :::|
  885 lLeuProLysLeuAspArgAspLeuLeu......SerSerTrpValAlaL 900

 5588 TG.........CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAA 5628
      ||           :::|||:::::      :::   |||
  900 euAlaGluThrThrThrArgGluValAlaValLeuAlaIleHisGlyAla 916

 5629 GGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGA 5678
      |||         |||::::::   ::::::   ::::::      :::
  917 GlyGlyAlaGlyLysSerArgAlaLeuGlnGluLeuLeuArgSerSerPr 933

 5679 CAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATT 5728
       :::|||      :::|||:::|||   :::        :::|||||||
  933 oGluLeu......AlaAspSerIleAsnIleValValProThrIleAsnL 948

 5729 TGATAGGTGCACTCAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTT 5778
      ||   :::      |||:::           |||      :::
  948 euAlaAsnAspTrpLysAlaLysLeuProGlnMetAspProArgArgVal 964

 5779 AAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGAC 5828
         :::|||:::      |||   :::                  |||::
  965 MetThrPheGlnLysAlaCysGluArgGlu............CysLysSe 977

 5829 GACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCA 5878
      ::::|||:::   |||::::::   |||   |||      ::::::   :
  977 rValThrIlePheAspAspTyrGlyLysLeuPro......AlaGlyPheV 992

 5879 TTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTG...... 5922
      ::   |||   ::::::   :::   :::::::::::|||:::
  992 alAspAlaTyrLeuAlaIleLysValAsnValGluLeuAlaIleLeuThr 1008

 5923 ..........................AAGAAGAGAGATCCGAATTTGGA 5945
                                 ::::::|||:::   ::::::::
 1009 GlyAspGlnArgGlnSerThrHisHisGlnGluArgGluSerGlnIleSe 1025

 5946 AGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCG 5995
      ::::|||:::|||      :::     |||:::   :::::::::   :
 1025 rSerLeuGlnSerAsnIleAlaGlnPheSerLysTyrAlaAspTyrTyrL 1042

 5996 TCAGGGGAATGTAT.................................. 6009
      ::::::::   :::
 1042 euAsnAlaThrHisArgGlnProArgArgLeuAlaAsnProIleLysVal 1058

 6010 ...............................AAAAGGGTTTTGGTGGA 6026
                                      |||      :::|||
 1059 HisAlaGluArgGlnLeuGlyGlyAlaValLeuLysAlaAsnIleValPr 1075

 6027 TGAGGTGCACATGATG.................CATCAAGGCTTACTA. 6057
       ::::::   |||:::                |||:::||||||
 1075 oAspLeuAlaMetValLeuValProAlaPheArgSerGlnSerLeuLeuT 1092

 6058 ..CAACTAGGCGTCTTCGCAACC......GGCGCGTCGGAAGGCCTCTTT 6099
        :::||||||      |||         ::::::   :::||||||
 1092 hrAspLeuGlyArgHisAlaMetThrTyrAlaGlyCysGlnGlyLeuThr 1108

 6100 TTTGGAGACATAAATCAGATA..............CCATTCATAAACAG 6134
      ::::::   ::::::   |||               |||:::   :::
 1109 LeuAsnHisLeuThrIleIleLeuAspLysAspThrProLeuCysSerAs 1125

 6135 GGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAA 6184
       |||   ::::::           :::::::|||          :::||||
 1125 pGluValLeuTyr............ThrAlaPheSerArgAlaSerGluS 1138

 6185 GCGTTGTATACACT..............TCTAAATCGTACAGGTGTCCG 6219
      ||:::::::::::              :::   ::::::   :::
 1138 erIleThrPheValAsnThrHisSerAspAsnProAlaPheLeuAlaLys 1154

 6220 TTAGATGTTTGCTACTTGTTGTCCTCAATGACC......GTAAGGGGAAC 6263
      ||||||:::         |||:::::::::         ||||||
 1155 LeuAspAlaThrProTyrLeuLysThrLeuIleSerTrpValArgGluAs 1171

 6264 GGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAG 6313
       |||:::                      :::|||   |||   |||:::
```

```
 1171 pGluGlu.....................AlaGlyAlaAspCysProAlaT 1181

 6314 TAAGATCGCTGTCCAAAAGG...CCAATTGGAACCACTGATGACGTAGCT 6360
      :::::   |||   |||      |||      |||         ||||||
 1181 hrGluProLeuValLysAspValProThrLysThrHisIleProValAla 1197

 6361 GAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
      :::   ::::::::::::         :::      :::|||   :::||
 1198 AsnAspLysValGlnLeuGluGlyLysIleGluAlaMetGluAspLysAs 1214

 6411 TATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAG 6460
      |   ::::::      :::|||:::   |||   :::   ::::::|||
 1214 pThrArgGluLeuTrpSerGlyGluGluLysThrAsnLeuMetGlnThrG 1231

 6461 TGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACG 6510
             :::                      |||   ||||||
 1231 lnAspProVal.....................ValGlnLeuPheProHis 1240

 6511 AAGAAAGCC...GATGACTCCCTATTCACTAAACAACCGCATATACTTGT 6557
      ::::::|||   |||::::::||||||
 1241 GlnGlnAlaLysAspGluAlaLeuPhe..................... 1249

 6558 TGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAA 6607
                                                      |
 1250 .....................................................L 1250

 6608 AGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCA 6657
      ||:::         :::|||      |||      |||:::|||::::::
 1250 ysIle......ThrIleGlyGluArgIleArgMetAlaThrProGluGln 1264

 6658 GTATCCGACGCTTTGCTTCACACG 6681
         :::      |||   ||||||
 1265 AsnAlaLysGlnLeuArgHisThr 1272

seq_name: pir2:JQ0096

seq_documentation_block:
hypothetical 176K protein - papaya mosaic virus
C;Species: papaya mosaic virus
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C;Accession: JQ0096
R;Sit, T.L.; Abouhaidar, M.G.; Holy, S.
J. Gen. Virol. 70, 2325-2331, 1989
A;Title: Nucleotide sequence of papaya mosaic virus RNA.
A;Reference number: JQ0096; MUID:89381685
A;Accession: JQ0096
A;Molecule type: genomic RNA
A;Residues: 1-1547 <SIT>
A;Cross-references: GB:D13957; GB:D00580; NID:g222330; PIDN:BAA03050.1; PID:g222331
A;Note: this protein shares two major domains of similarity with other potexvirus pro
C;Comment: The genome is a single-stranded, positive-sense RNA.
C;Superfamily: eggplant mosaic virus RNA-directed RNA polymerase


alignment_scores:
            Quality:  404.00            Length:    798
              Ratio:  1.101               Gaps:     35
 Percent Similarity:  45.990   Percent Identity: 23.559

alignment_block:
 US-09-301-906-3 x JQ0096   ..

 Align seg 1/1  to: JQ0096  from: 1  to: 1547

    4157 GTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGT 4206
         |||   |||   :::   :::|||      ::::::::::::      ||
     405 ValThrLeuThrTyrProThrLysThrPheAspAsnArgGlyTrpValVa 421

    4207 GAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACA 4256
         |:::   :::   :::::::|||   |||               :::||||
     421 lLysLeuGluAlaArgArgGlyTyrGluTrpPheAlaAspGluMetHisL 438

    4257 AACCGCGGTCAACGAAG.............................. 4273
```

```
     |||||::::::   :::
 438 ysProLysGlyProGluLeuAsnLeuGluGluLysLysThrAspProAsp 454

4274 ...GCGACAGGGAGCCAGTACAG.........TCCAGTCTTGTGAGTTC 4310
       |||:::   ::::::|||            |||   |||   :::::
 455 AlaAlaSerTyrGluLysTyrLeuLysAlaLeuSerLeuLeuGlnLysGl 471

4311 GCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGA 4360
     :|||::::::         |||   :::::::::      :::   |||:
 471 uProGluValMetGluAlaLysGluAlaGluAlaThrAspGluProGlnA 488

4361 AA...GAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACG 4407
     ::   |||||||||:::|||         |||::::::|||||   :::
 488 rgProGluValLysGluGluGlnAlaGluAlaSerThrSerGlyArgAla 504

4408 CCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAA 4457
           :::::::|||   ||||||   :::
 505 GluGluIleGlnGluAspProAlaThrLys..................... 514

4458 GATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGG 4507
                ||||||||
 515 ..........LysGlyLys............................... 517

4508 TACAAGTCGAGCAGCCCAAA...CAGAGGAGTTTGACGATCAATGAAGGC 4554
           |||:::|||:::   :::|||:::|||
 518 ........GluGluProAsnProAsnArgAspLeu.............. 526

4555 AAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCA 4604
                                     |||   ||||||:::::
 527 ............................LeuCysProCysGlyLeuHi 533

4605 GCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAAC.... 4650
     :|||   :::   |||             :::::::|||   :::
 533 sLeuLysIleLysAsn..............AlaGluPheProGluLeuP 545

4651 ..GCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTC 4698
       ::::::         |||:::|||   ||||||:::|||   ||||||
 545 roValLeuAspHisProAspHisLeuThrGlyArgLysAlaTrpPhePhe 561

4699 TCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATC 4748
     ||||||   |||:::   |||:::|||:::|||||||||||||:::|||
 562 SerLysAspGlyLysProTyrSerTyrThrGlyGlySerHisAlaSerAr 578

4749 AGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCA......ATTAAGT 4792
       |||||||||:::   ||||||   ||||||:::|||      ||||||
 578 gGlyTrpProAsnTrpLeuGluLysIleLeuAlaAlaIleGluIleLysG 595

4793 ACCCA...AGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGT 4839
       |||         |||::::::|||||||||||:::::::|||:::
 595 luProLeuProGluPheAsnGlnCysLeuValGlnGlnPheLysLeuGln 611

4840 GGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAA 4889
     :::::::::|||||||||   |||||||||   |||||||||:::   ::
 612 AlaAlaIleProPheHisArgAspAspGluProCysTyrProLysGlyHi 628

4890 CCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGT 4939
     :   :::||||||:::|||      |||::::::      :::      |
 628 sGlnValLeuThrIleAsnHisSerGlyGluCysLeuThrGlnIleAlaC 645

4940 GCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTAT 4989
     ||:::||||||               |||:::::::::   |||||||||
 645 ysGlnLysGly......LysAlaSerIleThrMetGlyPheGlyAspTyr 659

4990 TTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTC 5039
     :::|||   |||   |||||||||:::::|||   |||:::|||:::::
 660 TyrLeuSerProValGlyPheGlnGluSerHisLysHisAlaValSerAs 676

5040 CATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACT.......... 5079
     :         ||||||:::|||||||||||||||:::|||   .
 676 nThrThrGlyGlyArgValSerLeuThrPheArgCysThrValGlnGlnA 693

5079 ..................................................... 5079
```

```
 693 snLysPheAsnAspAspGlySerMetGluAlaLeuAspAsnLeuProTrp 709

5080 .....................CGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
                          :::::::::      ||||||:::||
  710 LysAlaTrpIleProLysLeuGlnAsnLeuGlyPheGlnGlyArgGlnLe 726

5109 GCAG..............:...TTAGCCGGCGGCGTGTCGGATGAGAAGT 5140
     ||||                |||   :::   :::::::::   :::|
  726 uGlnTyrAspProAsnGlyAlaLeuIleSerProIleGluGluIleArgS 743

5141 CACCAGGTGTTCCAAACCAGCAACCACAGAGC..................... 5172
     ||      :::|||:::   :::|||::::::
  743 er......MetProLysCysLysProGluGlyValProGluValValTyr 757

5173 .........CAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAA 5213
             :::|||   :::|||:::   ||||||   |||   :::
  758 LysThrLeuAspGlyLeuAlaArgAlaProThrProTyrSerProAsnPr 774

5214 GGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACT 5263
           :::::::::         :::||||||:::          :
  774 oIleArgAlaArgAlaTyrThrSerAspValLysAsnCysArgIleGlyA 791

5264 CGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313
     :::::      :::|||                  ||||||
  791 laLeuLeuArgGlnGlnGlyLys............GluTrp......... 800

5314 GATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGA 5363

  800 ..................................................... 800

5364 AGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGT 5413
                  |||         ||||||::||||||      |||
  801 .............GlyCysArgPheAspAlaLeuValGluAlaGlyLys. 812

5414 ATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGA 5463
                                                      |||
  813 .....................................................Arg 813

5464 AATATTGCCGTCACTACCGCCGAAGGTGTT...CTGAAAGTACCTAATCA 5510
     ::::::|||:::::::::   :::|||:::      :::      :::||
  814 GluLeuAlaIleSerValIleHisGlyAlaGlyGlySerGlyLysSerGl 830

5511 AGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATC 5560
     |:::   ::::::|||
  830 nAlaLeuGlnThrLeu................................. 835

5561 TCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTAC 5610
                                     ::::::|||   |||
  836 ..............................IleLysAspAsnPro... 840

5611 GTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGC 5660
                                                 ::::::|||
  841 .....................................GluLeuAspIl 844

5661 GGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGTTT 5710
      :::|||:::|||      |||
  844 eThrValValLeuProThrAsn.......................... 851

5711 TCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGC 5760
               ::::::|||            |||:::   ::::::
  852 ........GluLeuArgLeuAspTrpLeuArgLysLeuProLysAlaPro 865

5761 ATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCC 5810
     :::      |||:::||||||:::||||||   |||:::      |||||
  866 Gln......GluLysPheLysThrPheGlu...LysAlaLeuLeuAlaPr 879

5811 CCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGTCAC 5860
     ||||                 :::::   |||::::::   |||   |
  879 oPro.........ThrProIleValIlePheAspAspTyrGlyLysLeuP 893

5861 CCAAT............................AGCACAGCCACCATT 5880
     ||                                 ||||||:::   :::
  893 roAlaGlyTyrValGluAlaPheCysLeuTyrPheSerThrValGlnLeu 909
```

```
 5881 ACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAG 5930
               :::|||:::                                 |||::
  910 IleIleLeuThrGlyAspCys.....................LysGl 918

 5931 AGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGG 5980
      ::::   :::   |||:::   ::::::|||||||||   :::
  918 nSerValHisHisGluSerAsnGluAsnAlaThrThrSerSer..... 932

 5981 TGGTAAACTTTATCGTCAGGGGAATGTATAAA......AGGGTTTTGGTG 6024
        ::::::   :::|||:::         :::      |||      :::
  933 ..IleGluProLeuValLysGluAlaSerGluLeuCysArgTyrTyrIle 948

 6025 GATGAGGTGCACATGATGCATCAAGGCTTACTA...CAACTAGGCGTCTT 6071
      :::   :::|||         :::   |||      :::|||||||||::
  949 AsnAlaThrHisArgAsnLysLysAspLeuAlaAsnLysLeuGlyValTy 965

 6072 CGCA......ACCGGCGCGTCGGAA..................... 6090
      ::::      ||||||   :::|||
  965 rSerGluLysThrGlyLeuThrGluValThrHisGlyThrThrProIleP 982

 6091 ..GGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAG 6138
        ||||||   :::               :::|||   :::   ::::::
  982 roGlyLeuHisMet...........LeuValProSerLeuTyrLysLys 994

 6139 AAGGTGTTTAGG.......................ATGGATTGTGC 6161
      ::::::|||                                 |||
  995 GlnAlaPheSerGluMetGlyHisLysValSerThrTyrAlaGlyCysGl 1011

 6162 TGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACA 6211
         ::::::     |||   :::::::::   |||::::::::::
 1011 nGlyIleThrAlaProLysIleGlnIleLeuLeuThrGluGluThrSerL 1028

 6212 GGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTA 6255
        |||      :::|||   |||         |||   :::|||
 1028 euCysSerArgGluValLeuTyrThrAlaLeuSerArgAlaVal 1042
```

seq_name: pir2:S77910

seq_documentation_block:
hypothetical protein 1 - grapevine virus A
C;Species: grapevine virus A
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
C;Accession: S77910
R;Minafra, A.
submitted to the EMBL Data Library, June 1996
A;Reference number: S77910
A;Accession: S77910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1707 <MIN>
A;Cross-references: EMBL:X75433; NID:g1405614; PIDN:CAA53182.1; PID:g1405615
C;Superfamily: grapevine trichovirus B hypothetical protein

alignment_scores:
Quality: 394.00 Length: 1076
Ratio: 0.837 Gaps: 49
Percent Similarity: 43.773 Percent Identity: 21.654

alignment_block:
US-09-301-906-3 x S77910 ..

Align seg 1/1 to: S77910 from: 1 to: 1707

```
 3623 ATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGC 3672
      |||   |||:::||||||   ||||||||||::::::   :::   :::||
  383 IleAlaGlySerLeuPheThrArgLeuPheAsnArgProGluTyrAspAl 399

 3673 TCATTGTTAC...TGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGG 3719
      |:::|||        ::::::|||                        :
  399 aArgCysLeuGluLysPheIleArg.....................S 408
```

```
 3720 CTTGCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGG 3769
      ::|||   ||||||                          |||  ||||||
  408 erCysGluThrThrGluIle.................His..ValGluA 418

 3770 GGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTT..... 3813
        |||      |||   |||   :::|||         :::|||
  418 rgArg......TyrMetGluGlyIleArgArgGlyAlaSerPheLysVal 432

 3814 ............GACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCA 3851
                  |||:::   |||
  433 GlnAsnValMetAspTrpValGlu..................... 440

 3852 TGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCTAGTGTCTCAA 3901
                     ||||||:::   ::::::   |||:::|||:::
  441 ..............AspAspSerAlaAsnAlaLeuSerGluValAsnP 452

 3902 GCTATGATGTTGTCACAAATGTGCGCGACGTGGGGATTAGCACCAACGGG 3951
        |||                              ||||||   |||
  452 heLeuAsp.........................IleSerTrpAsnAsp 459

 3952 GAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACAC 4001
      :::|||:::         ||| |||      :::|||:::   :::
  460 ArgValSer.........Glu.ProTyrGlyIleGluCysIleHisGlyG 473

 4002 TTATGTCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAG 4051
        ||||||   |||:::|||      ::::::   |||      :::
  473 luGlySerArgIleArgValProLeuSerArgIleLeuArgAlaHisGlu 489

 4052 GTGATCCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGT 4101
      :::|||      |||   :::
  490 LeuIleAlaGlyValGlnThrAsp..................... 497

 4102 GTGAAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAAC 4151

  497 ..................................................  497

 4152 GGCTCGTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCG 4201
        |||:::::::::::            ||||||:::::|||
  498 .....ValGluIleAsnProProArgTyrValCysSerSerArgAlaLeu 512

 4202 TCGGTGAGGCGCCGACAC......AAGAAAGGGGAAGAGCTGCAGATGGT 4245
            |||::::::      |||   |||:::         |||
  513 IleHisPheArgGlnTyrLeuIleLysLeuGlyArgPheSerPheMet.. 528

 4246 AACACAGCACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTC 4295
             :::   |||:::   |||  |||   |||   :::||||::
  529 .........GluSerArgAlaIleLys..AspIleGluAspIleGlnAl 541

 4296 CAGTCTT............GTGAGTTCGCCACAGGCTGATATT....... 4326
      ::::|||           :::::::::   :::|||::::::
  541 aGlyLeuGluGluGlyValIleThrGluGluGluAlaGluLeuArgLeuL 558

 4327 ..............CCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAA 4362
                   ||||||:::|||      |||:::|||   :::
  558 euProThrThrLysProLysIleThr......GluIleHisMetAsp... 571

 4363 GAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAAT 4412

  571 ..................................................  571

 4413 CGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAA 4462
        |||:::   |||:::            |||   |||   :::   :
  572 ....AspAspThrProGly............ThrSerGlyGluSerAspV 583

 4463 TTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAA 4512
      :::::|||   |||:::|||
  583 alGluLysPheLysSerVal........................... 589

 4513 GTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGG 4562

  589 ..................................................  589

 4563 TAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATG 4612
```

```
      ::::::||||||                              :::   ::::
  590 .ArgSerLeuCys.........................ArgGluGluI 597

 4613 TGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTT 4662
      ::|||:::|||
  597 leTyrSerGlu..................................... 600

 4663 GTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGA 4712
         :::||||||||||||:::::::::|||:::|||:::   :::::
  601 ......LysLeuLysGlyArgGluValAlaPheTyrSerArgHisSerLy 615

 4713 GGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTG 4762
      :   |||   ||||||||||||||||||   |||   ||||||   :::|
  615 sGluTyrLysTyrAsnGlyGlySerHisArgSerLeuGlyTrpAspGluA 632

 4763 CCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCAC 4812
      |||||:::::::::         :::         :::   :::||||||
  632 laLeuAsnGluLeuThrGlnGluLeuGlyLeuAspAspSerTyrAspHis 648

 4813 TGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGA 4862
      ||||||:::|||:::|||      ||||||::::::   ||||||||||
  649 CysLeuIleGlnArgTyrThrAlaGlyGlySerIleGlyPheHisAlaAs 665

 4863 TGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCG 4912
      |||||||   ||||||         :::   ::::::||||||||||||
  665 pAspGluProCysTyrLeuProGlyGlySerValValThrValAsnLeuH 682

 4913 TGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATG 4962
        |||   |||:::|||:::::::|||   :::::::::||||||:::
  682 isGlyAspAlaThrPheGluValLysGluAsnGlnSerGlyLysIleGlu 698

 4963 GTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCA 5012
         ::::::   :::||||||   :::::::|||      |||:::||
  699 LysLysGluLeuHisAspGlyAspValTyrValMetGlyProGlyMetGl 715

 5013 AAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTT 5062
      |:::||||||   |||   |||:::|||      :::|||||   |||:
  715 nGlnThrHisLysHisArgValThrSerHisThrAspGlyArgCysSerI 732

 5063 TGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAG 5112
      ::|||:::|||                          ::::::   ::::::
  732 leThrLeuArg.......................AsnLysThrValAsp 740

 5113 TTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCA 5162
         :::      :::||||||   |||           ::::::
  741 TyrGluAlaArgLysGlyAspGluAspSerGluTyrGluGluAspLysAl 757

 5163 ACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCA 5212
         ::::::|||
  757 aGluLeuAspGluGlyIle.............................. 763

 5213 AGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATAC 5262

  763 ..................................................  763

 5263 TCGATATGGTGCGAACAAGATTACGTTAGGAAG.................. 5295
                      ||||||::::::|||
  764 ..................AspTyrLeuGlnLysAsnGlnGlyAsnMetCy 774

 5296 ..............TGTGAGTGGCTCAGGGCTGATAATCCA...GTGA 5326
                      ::::::   ::::::   ::::::|||   |||:
  774 sSerLeuLysAlaPheAlaAspHisMetGlnLeuSerThrProSerValI 791

 5327 TGGCTCTTGAACCTGACTACACCCCAATGACATTT...GAAGTGGTTAAA 5373
      ::|||:::               :::|||:::|||:::   |||:::
  791 leAlaIleValAsnGlyAlaSerProGlnThrLeuArgGluIleGluAsp 807

 5374 ACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTAT 5423
         |||   |||      ::::::|||:::      |||   |||   ::
  808 GlyGlyTyrSerLeuAlaThrLeuValAsnLeuSerLysAlaLeuAspPh 824

 5424 AGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCG 5473
      :                                              ||||||:
```

```
  824 e.....ProIleAlaI 828

 5474 TCACTACCGCCGAAGGTGTTCTGAAAGTACCT.........AATCAAGTT 5514
      ::         :::|||      ::::::|||         :::::::::
  828 leHisGlyGluArgGlyTyrAlaGluThrProGlySerTyrArgArgLeu 844

 5515 TATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCAT 5564
      :::   ::::::   :::   ||||||
  845 HisLeuLysIleThrSerGlyHisValGlu..................... 854

 5565 TTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGTAC 5614
                                                 |||:::
  855 .....................................ProPheGluG 858

 5615 TCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTG 5664
        :::::::::||||||   |||   :::                   :::
  858 lyValThrSerLysGlyGlyPheArgGlu.................Ala 868

 5665 GTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGTTTTCCA 5714
      :::   |||||||||::::::   |||
  869 MetLeuLeuGlyAspGlyValGlyVal...................... 877

 5715 CGATGCCATTAATTTGATAGGTGCACTGAAAGTC......GCTCGATGCG 5758
                          |||   :::::::|||      |||::::::
  878 ...................GlyHisPheArgValAspLysAlaLysAlaA 888

 5759 GCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCT 5808
           :::::::::||||||                          ||||||
  888 spArgLeuAlaGlnSerPhe.....................TyrAsn... 896

 5809 CCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTC........ 5850
                   |||:::|||   :::|||:::   ::::::
  897 ............GlyAsnThrGlyValLeuLeuGlyLysTyrAsnLysGl 909

 5851 ...................GTTAAGTCACCCAATAGCACAGCCACCATTA 5881
                          :::::::::|||:::::::::   |||
  909 yLysMetHisThrGlyGluIleGluGluProLysGluValLeuThrAlaP 926

 5882 CGGCTAATGTGGGAAGT.............................. 5898
        :::   :::||||||
  926 heGlyPheAlaGlySerGlyLysSerHisTrpCysGlnThrIleLeuLys 942

 5899 ......TCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTT 5942
            |||   :::      ::::::::::::   |||
  943 HisCysSerValGluLysValLeuValIleSerProArgLysValLeuAr 959

 5943 GGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTA 5992
        :::         :::      :::   :::   |||||||||:::|||
  959 gAspAspTrpValAlaLysIleSerLysLysHisArgValValThrPheG 976

 5993 TCGTCAGGGGAATG............TATAAAAGGGTTTTGGTGGATGAG 6030
        |||      |||              |||   :::::::::||||||
  976 luValAlaPheMetAspAspTyrGlyCysLysAspIleValIleAspGlu 992

 6031 GTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGG 6080
      :::   ::::::      |||   :::::::|||   :::   |||
  993 IleGlyLeuLeuProProGlyTyrIleAspLeuValIleAlaAlaHisGl 1009

 6081 CGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAA 6130
                   :::::::::||||||      |||      :::   :
 1009 nProArgThrLeuValLeuLeuGlyAspProLeuGlnSerThrTyrHisS 1026

 6131 ACAGGGAGAAGGTGTTTAGGATGGATTGTGCT.........GTTTTTGTT 6171
      ::::::::              ::::::::::::          ||||||
 1026 erLysArgAspAsnValValLeuGluAlaSerGlnGluAspValPheAsn 1042

 6172 CCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTT 6221
         :::   ::::::   |||         |||:::|||   |||
 1043 ArgValArgGlyLysLeuProTyrLeuCysTyrSerHisArgLeuProAr 1059

 6222 AGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGT 6271
        :::   |||   |||:::   ::::::          |||:::|||:::
 1059 gAsn...CysLysLeuPhe...GluIleGluCysMetGlyAlaGluSer. 1073
```

```
    6272 GTTACCCTGAAAAGGTCGTTAGCGGT...........AAGGACAAACCA 6309
                 ::::::||||||                 ||||||:::|||
    1074 .....GluLysArgValValTyrArgSerAsnArgLeuLysAspGluPro 1088

    6310 GTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGC 6359
         ::::::
    1089 ThrIle..................................................... 1090

    6360 TGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGG 6409
                                 |||   |||:::      |||::::
    1091 ........................CysAlaThrArgAlaMetLysGluG 1099

    6410 ATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACA 6459
         ::                |||||||:::|||                 |||
    1099 lu..............LysGlySerGly...........TrpTyrThr 1106

    6460 GTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGG.. 6507
         |||   |||:::||||||   :::|||::::::   :::::::::
    1107 ValSerGluThrGlnGlyLeuSerPheLysSerCysLeuIleTyrLeuAs 1123

    6508 ..........ACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGC 6547
                     :::||||||   |||::::::
    1123 pGluHisTrpAlaLysLysGluAspGluAsp..................... 1133

    6548 ATATACTTGTTGGTTTGTCGAGACACACACGCTCACTG.........GTT 6588
           ::::::|||:::|||:::|||         ::::::         |||
    1134 ..ValMetValAlaLeuThrArgSerArgGlyGluIleGlyIleHisVal 1149

    6589 TATGCCGCTCTGAGCTCAAAGTTG 6612
               ||||||::::::||||||
    1150 ThrProAlaLeuLysLysLysLeu 1157

seq_name: pir2:S77908

seq_documentation_block:
hypothetical protein 1 - grapevine trichovirus B
C;Species: grapevine trichovirus B
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
C;Accession: S77908
R;Saldarelli, P.
submitted to the EMBL Data Library, March 1996
A;Reference number: S77908
A;Accession: S77908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1707 <SAL>
A;Cross-references: EMBL:X75448; NID:g1220265; PIDN:CAA53196.1; PID:g1061119
C;Superfamily: grapevine trichovirus B hypothetical protein


alignment_scores:
            Quality:  391.50              Length:   1083
              Ratio:  0.884                Gaps:     42
 Percent Similarity:  40.905   Percent Identity: 20.222

alignment_block:
 US-09-301-906-3 x S77908   ..

 Align seg 1/1  to: S77908  from: 1  to: 1707

    3943 ACCAACGGGGAAGTTACTGGTGAAGAAGAG.........ACCCATTCACC 3983
         ||||||||||||   ::      |||            ::::::
     338 ThrAsnGlyGluGluSerThrAlaGluAlaLeuPheValAlaGlnLeuSe 354

    3984 TCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCT.... 4029
          |||:::::::|||   |||   :::   ::::::::::::
     354 rArgGlnIleGlnGluThrLysLeuTyrAspLysMetGlyAsnPheSerL 371

    4030 .................GCAGCTGTGGCGGAAAGACAAGGTGATCCGTCG 4062
                           :::||||||::::::   :::||||||
     371 euLysGluCysIleTrpSerAlaValSerSerAlaMetGlyAspSerLeu 387

    4063 GGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGG 4112
                              |||      :::   ||||||:::::::::
     388 ValTyrPhePheAspLysAlaArgPheTyrAsnGluSerLeuGluSerPh 404

    4113 TGTC.....................................................G 4117
          :::                                                    |
     404 eIleLeuAsnCysGlyProAlaLysIleGluIleLysArgValTyrArgA 421

    4118 ACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAG 4167
         ||   |||::::::::::|||      |||          :::   |||:::
     421 spAlaValTyrArgAsnGln......GlyValCysProHisLeuValSer 435

    4168 GTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGAC 4217
            ||||||:::         |||:::::::|||:::   |||
     436 PheAspGlySer.........ProAspGlyValMetAspGluLeuTyrLe 449

    4218 ACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCA 4267
          |||:::                                          ::::
     449 uGlnHis.....................................IleA 453

    4268 ACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAG 4317
         :::::   ||||||:::              ||||||:::|||:::
     453 rgAspMetAspArgSerSerGluProTyrSerLeuThrSerAla...... 467

    4318 GCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGT 4367
                    :::|||   :::|||::::::   ::::::
     468 ...........GluValLeuGluSerSerIleGlyCysArg........ 477

    4368 GAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCG 4417

     477 ..................................................... 477

    4418 ATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGAC 4467
                                                 :::::::::|||:::
     478 ..............................IleSerLeuIleGlu 482

    4468 AAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGA 4517
         |||:::|||:::           :::::::::::|||:::::::
     483 MetGlyArgAlaMetLeuLysPheHisGlnGlnGlnAspValMetIleAs 499

    4518 GCAG.............................CCCAAACAGAGGA 4534
         ::::                                ::::::
     499 nGluArgValValLeuSerProLeuAspAspPheArgSerMetArgLysA 516

    4535 GTTTGACGATCAATGAAGGCAAGGCCGGT..................... 4563
         ::||||||   ::::::|||      |||
     516 laLeuThrAspArgAspGlyIleLeuGlyThrThrCysLeuLeuLeuGlu 532

    4564 .................AAACAGCTTTGCATG................... 4578
                           |||:::      |||
     533 TrpGluTyrAspProLeuLysArgTyrSerMetGlyLeuIleThrAlaTh 549

    4578 ..................................................... 4578

     549 rGluLeuAsnAlaAlaIleIleGlyThrThrLysIleThrAlaLysLysG 566

    4579 .......................TTTAGAACGTGTTCCTGCGGTGTG 4602
                                 ::::::   ||||||||||||||||
     566 luMetAlaLysGluCysSerGluArgTyrGlnGlnCysSerCysGlyVal 582

    4603 CAGCTGGATGTGTACAACGAAGCGACT.........ATCGCCACCAGGTT 4643
         :::|||   :::   :::      |||          :::   |||:::
     583 ArgLeuProMetArgLysLeuGluThrProLeuLysValProThrGlu.. 598

    4644 CTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCT 4693
                        |||:::|||   ||||||||||||   |||:::|
     599 .............GluPheThrAspAlaLeuLysGlyArgArgAlaAlaP 611

    4694 TTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTT 4743
         ||:::|||:::   :::   :::|||:::|||:::|||::::::|||:::
     611 heTyrSerArgHisSerTrpAsnTyrSerTyrThrGlyAlaAsnHisAla 627

    4744 TCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTA 4793
```

```
    |||:::|||||||||:::   ::::::::::|||         :::
628 SerLysGlyTrpProGluTrpValSerAsnIleAlaGluLysLeuGlyLe 644

4794 CCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAG 4843
       :::   ||||||||||||||||:::|||   |||:::         |
 644 uGlyGluAspPheAspHisCysLeuAlaGlnIleTyrGluGluAspLysG 661

4844 GCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCT 4893
     ||:::   |||:::||||||||||||   ||||||   :::|||
 661 lyIleAsnPheAsnAlaAspAspGluProCysTyr...ThrAspProGlu 676

4894 ATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAG 4943
     ::::::||||||||||||   |||:::|||||||||      ||||||
 677 ValValThrValAsnLeuAsnGlyAsnAlaAsnPheHisLeuLysCys.. 692

4944 GAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTC 4993
            |||:::      |||:::   :::::::::||||||   ::::
 693 .......GlySerGluSerValValProLeuSerAspGlyAspValLeuI 707

4994 TTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATC 5043
     ::||||||   |||||||||:::||||||   |||:::|||::::::
 707 leMetProLysGlyPheGlnLysThrHisLysHisAlaValThrGlyThr 723

5044 GACGAAGGGCGCATCAGTTTGACGTTCAGG..................... 5073
     :::   ||||||||||||||||||||||
 724 SerAlaGlyArgIleSerLeuThrPheArgAsnGlyIleAsnAlaProAs 740

5074 ............................GCAACTCGGCGCGTCTTTG 5092
                                    |||:::|||:::   |
 740 pGluGlySerAspLysMetSerGluTyrGluGluThrHisArgIleGlyG 757

5093 GTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142
     ||                                          :::
 757 ly.........................................Lys 758

5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
     ||||||   |||:::::::::   :::::::::   :::   :::   ::
 759 ProGlyCysProArgGluGluGlnLysGluSerValSerLeuGlnIleLe 775

5193 CACACCA..................................... 5199
     ::::
 775 uAlaAspHisMetArgValAspLeuAlaIleCysThrSerMetValPheA 792

5200 ..........................AAATCGGGGGGC......... 5211
                                 |||   ||||||
 792 laLysAspProArgAlaArgGluGluValLysArgGlyGlyMetThrLeu 808

5212 ....................AAGGCTCTATCTGAGGGAAGTGGTAGGGA 5240
                         |||:::|||:::   |||:::      ||
 809 GlyThrPheIleCysValLeuLysSerLeuAsnLeuGlyAlaTyrIleGl 825

5241 AGTCAAGGGGAGGTCGACATAC.......................TCGA 5266
     |   :::            |||                       |||
 825 uSerGluArgGlyCysLeuTyrValAsnGlyAlaTyrLysGluLeuSerC 842

5267 TATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGAT 5316
       ::::::|||:::|||::::::   :::   :::      |||   :::
 842 ysTyrAlaGluAspAspHisIleSerGluTrpSerGlyProArgLysGlu 858

5317 .........AATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGAC 5357
             :::   ::::::::::::::::||||||         :::
 859 ThrThrPheSerAsnAlaLeuSerMetAsnProAspIleArgValValLy 875

5358 ATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACT 5407
      :::||||||   :::
 875 sTyrGluValCysGluGln............................. 881

5408 TGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATG 5457

 882 .............................................Arg 882

5458 GCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAA 5507
     |||::::::::::::   :::      ||||||         |||
```

```
 883 AlaLysLysLeuValAspSerPheGlnGluGlyPheThrGlyVal..... 897

5508 TCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAG 5557
                    |||      |||:::      ||||||:::   :
 898 .............CysLeuAsnLysPheGlnLysGlnLysSerSerPheA 910

5558 ATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCG 5607
     ::|||:::                 :::|||   ::::::   |||:::
 910 snLeuVal...............AsnGlyLysGluLeuIleAspValHis 921

5608 TACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGA 5657
        :::|||           |||:::      :::||||||:::
 922 LeuThrLeu............GlyPheAlaGlySerGlyLysSerPheTy 934

5658 CGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGG 5707
             :::|||   ::::::
 934 rProGlnCysValLeuLysAsnSer........................ 942

5708 TTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGC 5757
          |||   |||   :::|||:::   :::   :::   |||   |||
 943 .....HisTyrAlaAsnSerLeuValIleValProArgLysAlaLeuCys 957

5758 GGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGC 5807
      :::   |||:::|||::::::   |||
 958 .SerAspTrpSerAsnLysValHisPro..................... 966

5808 TCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGT 5857

 966 ............................................... 966

5858 CACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGAC 5907

 966 ............................................... 966

5908 ATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAG 5957
                          :::::::::::          :::::::
 967 ......................AspValLysValValThrPheGluSerA 976

5958 TGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGT 6007
     :::::   |||          |||:::  ::::::||||||
 976 laPheArgGlnGlnLysLysGlyTyr..GlyLeuIleVal.......... 988

6008 ATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTA 6057
                    :::||||||:::   ::::::      |||   :::
 989 ..............IleAspGluIleGlyLeuLeuProProGlyTyrIle 1000

6058 CAACTA......GGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTT 6101
     ::::::       |||   |||:::      :::          |||   ::
1001 AspMetValHisGlyTyrPheCysTyrAspSerLeu......LeuValLe 1015

6102 TGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGA 6151
     :||||||      |||      :::   ::::::          |||
1015 uGlyAspProLeuGlnCysGluTyrHisSerLysAlaAspHisPhe.... 1030

6152 TGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTA......... 6192
                    |||:::   ::::::|||||||||
1031 ..............PheLeuGlyGlnGluGluSerValPheLysLysPhe 1042

6193 ..............TACACTTCTAAATCGTACAGGTGTCCGTTAGATGT 6227
                     |||      ||||||:::|||   |||   :::
1043 LysGlyHisCysAsnTyrLeuTyrLysSerHisArgLeuProArgAsn.. 1058

6228 TTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACC 6277
               |||   ::::::          |||:::|||
1059 ....GlnLysLeuPheGluIleGluCysAspGlyAlaGlu.........G 1071

6278 CTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCC 6327
       |||   |||:::   :::|||   :::
1071 lyGluGlyValThrPheAsnLysProArgAla.................. 1081

6328 AAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGT 6377
                                                  |||::
1082 .....................................LysAspLe 1084
```

```
    6378 GTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGA 6427
         :   ||||||   :::|||                                :
    1084 uThrLeuCysAlaSerGln............................A 1091

    6428 AGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGA 6477
         ::   |||   :::|||      |||   ||||||   |||:::||||||
    1091 rgArgLysAspSerGluLysAspValAsnThrValGlyGluSerGlnGly 1107

    6478 AAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTC 6527
            :::   :::   |||   ::::::      |||      |||   ||
    1108 LeuSerAlaAsnArgValAsnIleLeuLeuAspLys......AspTrpSe 1122

    6528 CCTA..........................TTCACT.........AAAC 6541
         ||||                         ||||||         |||:
    1122 rLeuValAsnAspGluThrValIleValAlaPheThrArgAlaArgLysG 1139

    6542 AACCGCATATA..................................... 6552
         ::   :::|||
    1139 luIleAsnIleIleGlyAspAlaSerLeuValAsnAsnLeuLysArgSer 1155

    6553 ......................CTTGTTGGTTTGTCGAGACACACACG 6578
                                 :::::::|||   :::|||   |||::
    1156 AlaLysSerThrValLeuLysLysIleLeuGlyGlyGluArgValThrGl 1172

    6579 CTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTC 6624
         ::::|||:::   :::   :::   :::||||||   |||   :::
    1172 uGlyLeuIleLeuSerLeuIleArgLysLysLeuProAspValIle 1187
```

seq_name: pir2:S01865

```
seq_documentation_block:
genome polyprotein - tobacco rattle virus (strain SYM)
N;Alternate names: RNA replicase
N;Contains: 134K protein; 194K protein; RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: tobacco rattle virus, TRV
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 17-Mar-2000
C;Accession: S01865
R;Hamilton, W.D.O.; Boccara, M.; Robinson, D.J.; Baulcombe, D.C.
J. Gen. Virol. 68, 2563-2575, 1987
A;Title: The complete nucleotide sequence of tobacco rattle virus RNA-1.
A;Reference number: S01865; MUID:88034943
A;Accession: S01865
A;Molecule type: genomic RNA
A;Residues: 1-1707 <HAM>
A;Cross-references: EMBL:X06172
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: nucleotidyltransferase
F;1-1707/Product: 194K protein #status predicted <MAT2>
F;1-1187/Product: 134K protein #status predicted <MAT1>
F;1188/Region: opal stop codon


alignment_scores:
            Quality:  338.50              Length:   1858
              Ratio:  0.486                Gaps:     72
 Percent Similarity:  37.513   Percent Identity: 17.976

alignment_block:
 US-09-301-906-3 x S01865   ..

 Align seg 1/1  to: S01865  from: 1  to: 1707

    1333 GACGCAATGATTCGCGAAAAACCCTCA.................CATAG 1364
         |||   ||||||||||||:::   :::                 |||::
      50 AspAspMetIleArgGluArgLysThrLysAspProValIleValHisGl 66

    1365 GTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCT 1414
         :            |||:::   :::|||::::::||||||   |||::::
      66 u.........ValLeuSerGlnLysGluGlnAsnLysLeuMetGluIleT 80

    1415 TTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCA 1464
         ::||||||::::::|||   |||:::|||:::   :::      |||
```

```
      80 yrProGluPheAsnIleValPheLysAspAspLysAsnMetValHisGly 96

    1465 TTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTG 1514
         ||||||   |||   |||:::             ::::::      |||
      97 PheAlaAlaAlaGluArgLysLeuGlnAlaLeuLeuLeuLeuAspArgVa 113

    1515 TGGTAATGTGTGC...TTCTTCGATATTGGGGGGAGCTTCACGTATCATG 1561
                :::           ||||||||||||:::::::::::   |
     113 lProAlaLeuGlnGluValAspAspIleGlyGlyGlnTrpSerPheTrpV 130

    1562 TCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTT 1611
         ||      |||:::   :::   |||   |||   |||   |||||||:::
     130 alThrArgGlyGluLysArgIleHisSerCysCysProAsnLeuAspIle 146

    1612 AAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGG 1661
         :::|||   :::|||:::|||::::::   :::||||||::::::   ||
     147 ArgAspAspGlnArgGluIleSerArgGlnIlePheLeuThrAlaIleGl 163

    1662 GGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAA............. 1698
         |   ::::::      ::::::      ::::::|||
     163 yAspGlnAlaArgSerGlyLysArgGlnMetSerGluAsnGluLeuTrpM 180

    1699 ......................GCGGCTTCAAAGTCTGTGTCTTACTGT 1725
                                 |||:::   ::::::      |||
     180 etTyrAspGlnPheArgLysAsnIleAlaAlaProAsnAlaValArgCys 196

    1726 AGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCCGGGTTTT..... 1770
         ::::::      |||:::|||      |||         ||||||
     197 AsnAsnThrTyrGlnGlyCysThrCysArg.........GlyPheSerAs 210

    1771 ................................ATGGTGGATGTGTACG 1786
                                         :::   :::::||||
     210 pGlyLysLysLysGlyAlaGlnTyrAlaIleAlaLeuHisSerLeuTyrA 227

    1787 ATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTG 1836
         ||::::::   :::::::::      :::|||   :::|||   :::
     227 spPheLysLeuLysAspLeuMetAlaThrMetValGluLysLysThrLys 243

    1837 GTTTTCGACATAGCTCTTATGTTC...CCCGTGGAGTTGTTGTACGGTAA 1883
         |||        |||::::::|||   |||   ::::::|||      ::
     244 ValValHisAlaAlaMetLeuPheAlaProGluSerMetLeuValAspGl 260

    1884 CGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAGGTG 1933
         :|||                              |||   :::   :::||
     260 uGly.......................ProLeuProSerValAspGlyT 269

    1934 ATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAACATTCC 1983
           |||:::      |||            |||:::::::|||
     269 yrTyrMetLysLysAsn............GlyLysIleTyr......... 278

    1984 TTCTCTAACGTAAGCGGGTTTTTC.........ACCTTTTCTTATGTACG 2024
                       ||||||            :::|||||||||:::::
     279 ............PheGlyPheGluLysAspProSerPheSerTyrIleHi 291

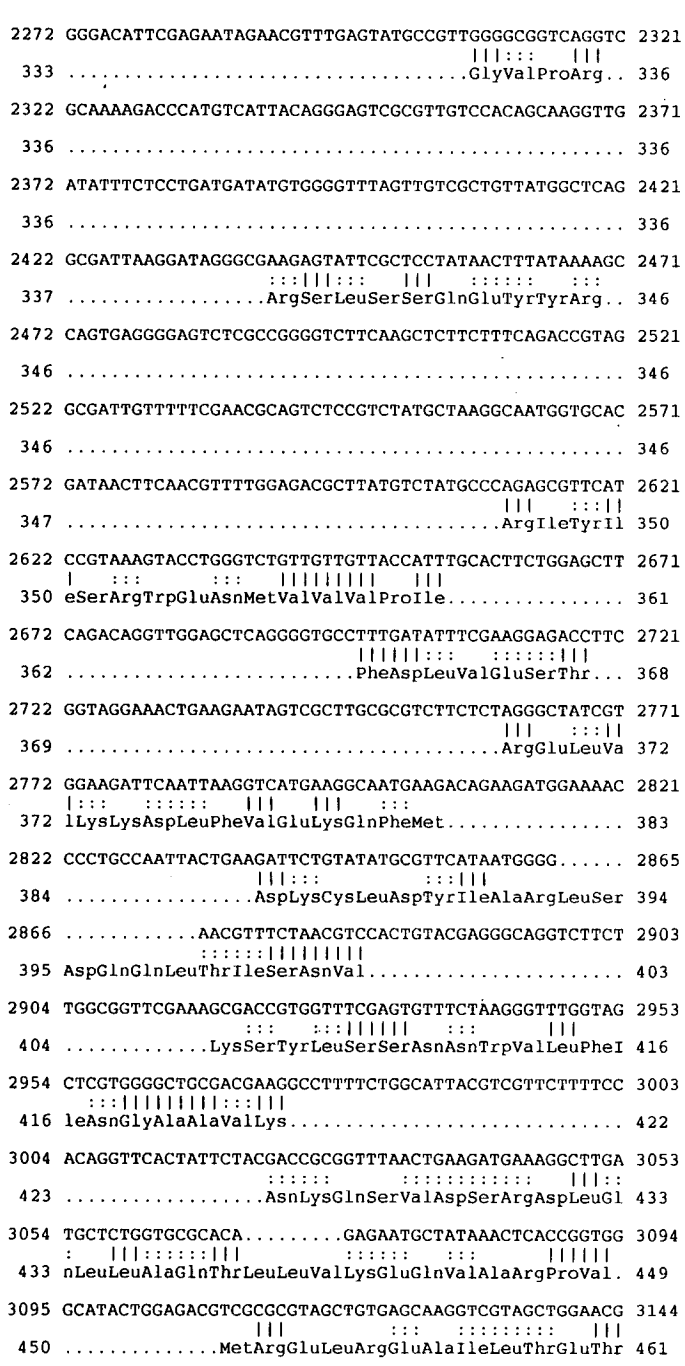
    2025 CACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTG 2074
         :      :::                  ||||||:::   |||
     291 sAspTrpGlu.................GluTyrLysLysTyrLeuLeu. 301

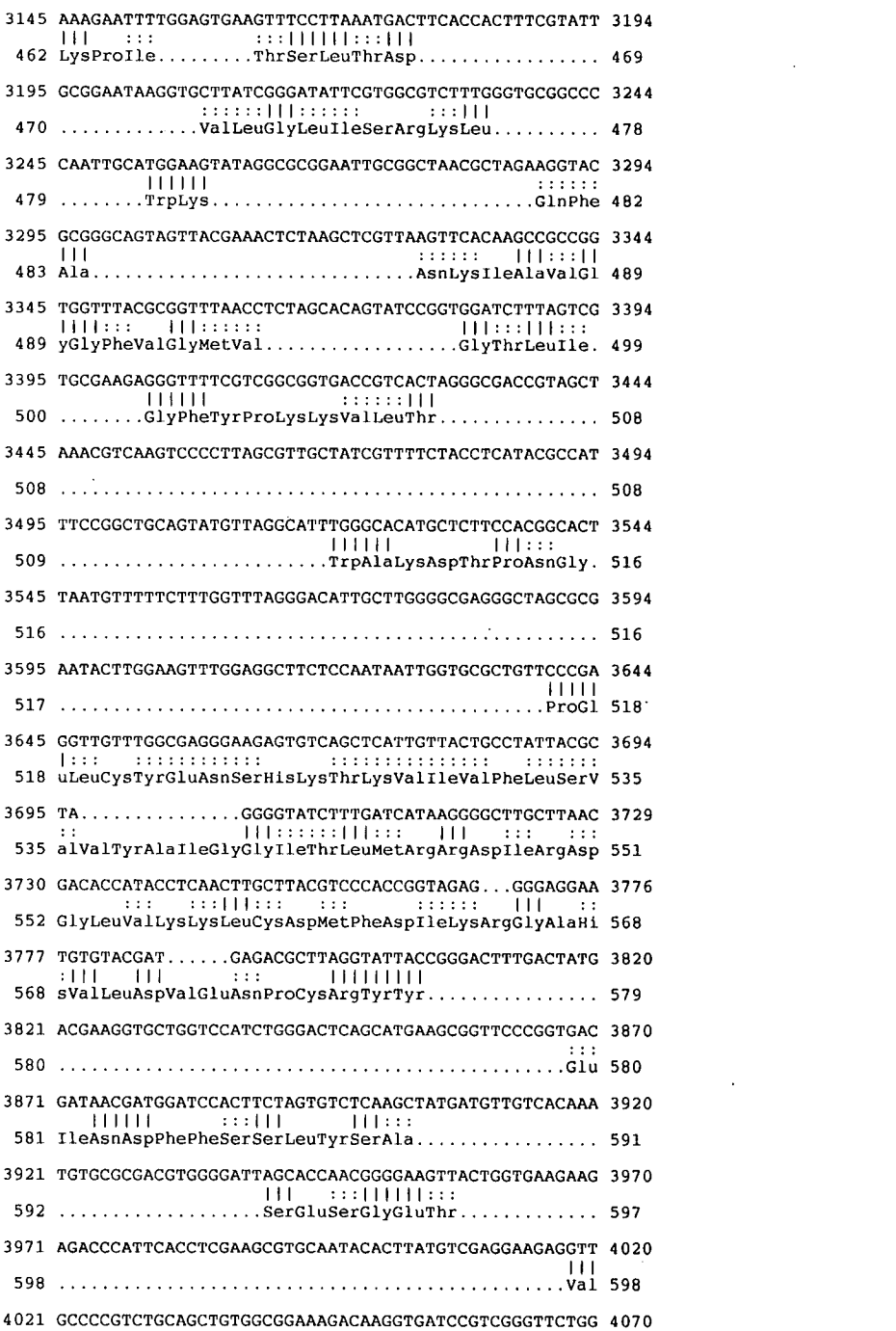
    2075 GTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACTGAG 2124
                                                  |||
     302 .....................................GlyLysPro 304

    2125 GTTACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTT 2174
         |||:::|||:::                     |||:::   :::   ||
     305 ValSerTyrGln.....................GlyAsnValPheTyrPh 314

    2175 CATACCT...GTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCC 2221
         |   |||      |||   |||:::::::::   |||:::
     314 eGluProTrpGlnValArgGlyAspThrMetLeuPheSer.......... 327

    2222 TCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATA 2271
                              ||||||      |||
     328 ......................IleTyrArgIleAla............. 332
```


2272 GGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTC 2321
333 GlyValProArg.. 336
2322 GCAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTG 2371
336 336
2372 ATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAG 2421
336 336
2422 GCGATTAAGGATAGGGCGAAGAGTATTCGCTCCTATAACTTTATAAAAGC 2471
337 ArgSerLeuSerSerGlnGluTyrTyrArg.. 346
2472 CAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAG 2521
346 346
2522 GCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCAC 2571
346 346
2572 GATAACTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCAT 2621
347 ArgIleTyrIl 350
2622 CCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTT 2671
350 eSerArgTrpGluAsnMetValValValProIle..... 361
2672 CAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTC 2721
362 PheAspLeuValGluSerThr... 368
2722 GGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGT 2771
369 ArgGluLeuVa 372
2772 GGAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAAC 2821
372 lLysLysAspLeuPheValGluLysGlnPheMet..... 383
2822 CCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGG..... 2865
384 AspLysCysLeuAspTyrIleAlaArgLeuSer 394
2866 AACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCT 2903
395 AspGlnGlnLeuThrIleSerAsnVal..... 403
2904 TGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAG 2953
404 LysSerTyrLeuSerSerAsnAsnTrpValLeuPheI 416
2954 CTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCC 3003
416 leAsnGlyAlaAlaValLys..... 422
3004 ACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGA 3053
423 AsnLysGlnSerValAspSerArgAspLeuGl 433
3054 TGCTCTGGTGCGCACA.....GAGAATGCTATAAACTCACCGGTGG 3094
433 nLeuLeuAlaGlnThrLeuLeuValLysGluGlnValAlaArgProVal. 449
3095 GCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACG 3144
450 MetArgGluLeuArgGluAlaIleLeuThrGluThr 461


3145 AAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATT 3194
462 LysProIle.....ThrSerLeuThrAsp..... 469
3195 GCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCC 3244
470 ValLeuGlyLeuIleSerArgLysLeu..... 478
3245 CAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTAC 3294
479 TrpLys.....GlnPhe 482
3295 GCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGG 3344
483 Ala.....AsnLysIleAlaValGl 489
3345 TGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCG 3394
489 yGlyPheValGlyMetVal.....GlyThrLeuIle. 499
3395 TGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCT 3444
500 GlyPheTyrProLysLysValLeuThr..... 508
3445 AAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCAT 3494
508 508
3495 TTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACT 3544
509 TrpAlaLysAspThrProAsnGly. 516
3545 TAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCG 3594
516 516
3595 AATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGA 3644
517 ProGl 518
3645 GGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGC 3694
518 uLeuCysTyrGluAsnSerHisLysThrLysValIleValPheLeuSerV 535
3695 TA.....GGGGTATCTTTGATCATAAGGGGCTTGCTTAAC 3729
535 alValTyrAlaIleGlyGlyIleThrLeuMetArgArgAspIleArgAsp 551
3730 GACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAG...GGGAGGAA 3776
552 GlyLeuValLysLysLeuCysAspMetPheAspIleLysArgGlyAlaHi 568
3777 TGTGTACGAT.....GAGACGCTTAGGTATTACCGGGACTTTGACTATG 3820
568 sValLeuAspValGluAsnProCysArgTyrTyr..... 579
3821 ACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGAC 3870
580 Glu 580
3871 GATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAA 3920
581 IleAsnAspPhePheSerSerLeuTyrSerAla..... 591
3921 TGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAG 3970
592 SerGluSerGlyGluThr..... 597
3971 AGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTT 4020
598 Val 598
4021 GCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGG 4070

```
      |||:::   :::   ::::::   :::::::|||
  599 LeuProAspLeuSerGluValLysAlaLysSerAsp..............  610

 4071 TACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACG 4120
                          ::::::::::::   |||:::   :::|||:
  611 ................LysLeuLeuGlnGlnLysLysGluIleAlaAspG  622

 4121 ATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTG 4170
      ::   :::       :::   |||:::   ::::::   :::|||::::::
  622 luPheLeuSerAlaLysPheSerAsnTyrSerGlySerSerValArgThr  638

 4171 GACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACA 4220
      :::                 |||   ||||||||||||::::::   :::
  639 Ser..............ProProSerValValGlySerSerArgSerGl  650

 4221 AGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACG 4270
                          |||:::|||:::
  650 yLeuGlyLeuLeuLeuGluAspSerAsnVal..................  660

 4271 AAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCT 4320

  660 .................................................  660

 4321 GATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTG.. 4368
                  :::||||||::::::|||   ::::::::::::|||
  661 ............LeuThrGlnAlaArgValGlyValSerArgLysValAs  673

 4369 .AAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCG 4417
          :::|||:::           :::||||||
  673 pAspGluGluIleMetGluGlnPheLeuSerGly..............  684

 4418 ATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGAC 4467
                                                 :::||||||
  685 .....................................LeuIleAsp  687

 4468 AAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGA 4517
                  :::   :::|||::::::      :::::::::   ||
  688 ThrGluAlaGluIleAspGluValValSerAlaPheSerAlaGluCysGl  704

 4518 GCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAAC 4567
      |:::                        :::|||   ::::::   |||
  704 uArg.....................GlyGluThrSerGlyThrLysV  713

 4568 AGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTAC 4617
        ||||||         |||
  713 alLeuCysLysProLeuThr.........................  719

 4618 AACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGA 4667

  719 .................................................  719

 4668 TAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGT 4717

  719 .................................................  719

 4718 ATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTA 4767
                                              |||   ::::::
  720 ...................................ProProGlyPhe  723

 4768 GAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTT 4817
      |||::::::|||   |||:::|||                       ||
  724 GluAsnValLeuProAlaValLys.....................ProLe  733

 4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACG 4867
      ||||:::|||   |||
  733 uValSerLysGlyLys..............................  738

 4868 AGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGG 4917
                                       ||||||:::   |||
  739 ............................ThrValLysArgVal...  743

 4918 AAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCAT 4967
```

```
  743 .................................................  743

 4968 AAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTT.......  5010
                      |||||||||   :::       |||
  744 ...............AspTyrPheGlnValMetGlyGlyGluArgLeuP  755

 5011 ..CAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATC 5058
        :::|||      :::   :::   ||||||:::|||      |||
  755 roLysArgProValValSerGlyAspAsnSerValAspAlaArgArgGlu  771

 5059 AGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
         |||   :::      ::::::||||||            ::::::::
  772 PheLeuTyrTyrLeuAspAlaGluArgValAlaGlnAsnAspGluIleMe  788

 5109 GCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACC 5158
      ::::|||
  788 tSerLeu.....................................T  791

 5159 AGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGG 5208
        :::      |||:::|||:::   ||||||               |||
  791 yrArgAspTyrSerArgGlyValIleArgThr.............Gly  802

 5209 GGCAAGGCTCTATCTGAGGGAAGTGGT.........AGGGAAGTCAAGGG 5249
      |||:::         :::|||   |||            |||:::||||::
  803 GlyGlnAsnTyrProHisGlyLeuGlyValTrpAspValGluMetLysAs  819

 5250 GAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTG 5299
      :                  ||||||   :::      |||   :::
  819 n.................TrpCysIleArgProValValThrGluHisA  830

 5300 AGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACC 5349
        ::::::                        ::::::||||||
  830 laTyrVal.....................PheGlnProAsp......  836

 5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
                                                    :::
  837 ...................................LysArgMetAs  840

 5400 GGAGTACTTGAAGTATCTGGCTATAGGCATT...GAGAGGACATACAGGG 5446
       ::::::      ||||||   :::::::::   ||||||
  840 pAspTrpSerGlyTyrLeuGluValAlaValTrpGluArg.......  853

 5447 CGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTG 5496

  853 .................................................  853

 5497 AAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAA 5546

  853 .................................................  853

 5547 GTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGA 5596
                                          |||:::   |||:
  854 ...............................GlyMetLeuValA  858

 5597 GAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGC 5646
      ::|||:::
  858 snAspPhe.....................................  860

 5647 AAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGA 5696
                |||||      :::::::|||   ::::::||||||||
  861 ...........AlaValGluArgMetSerAspTyrValIleValCysAs  873

 5697 CGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAG 5746
      |:::      :::   ::::::   :::   ::::::      |||::::
  873 pGlnThrTyrLeuCysAsnAsnArgLeu...IleLeuAspAsnLeuSerA  889

 5747 TCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAG 5796
      ::         |||   |||:::   ||||||               :::
  889 laLeuAspLeuGlyProValAsnCysSerPhe.............Glu  900

 5797 TGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGA 5846
            ::::::   ||||||   ||||||:::|||   :::|||:::
  901 LeuValAspGlyValProGlyCysGlyLysSerThrMetIleValAsn..  916
```

```
5847 ATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA. 5895
             |||   |||      :::   ::::::   ::::::|||
 917 ..........SerAlaAsnProCysValAspValValLeuSerThrGlyA 930

5896 ..............................AGTTCTGAGGACATAAAT 5913
                                   :::|||:::   :::
 930 rgAlaAlaThrAspAspLeuIleGluArgPheAlaSerLysGlyPhePro 946

5914 ATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTAC 5963
        :::|||:::|||                          :::
 947 CysLysLeuLysArgArg..........................ValLy 954

5964 CACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAA 6013
      ||||||:::|||   :::::::::   :::      ::::::
 954 sThrValAspSerPheLeuMetHisCysValAspGlySerLeuThrGlyA 971

6014 GGGTTTTG...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAA 6060
       ||||||      ||||||:::   |||   |||   |||::::::
 971 spValLeuHisPheAspGluAlaLeuMetAlaHisAlaGlyMetValTyr 987

6061 CTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACAT 6110
     :::   :::      :::||||||::::::   :::      ||||||
 988 PheCysAlaGlnIleAlaGlyAlaLysArgCysIleCysGlnGlyAspGl 1004

6111 AAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTG 6160
      |||||||||   |||      |||   :::         :::      :
1004 nAsnGlnIleSerPheLysProArgValSerGlnValAspLeuArgPheS 1021

6161 CTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTAC 6210
     ::   :::|||   |||   :::   |||:::         ::::::|||
1021 erSerLeuValGlyLysPheAspIleValThrGluLysArgGluThrTyr 1037

6211 AGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGG 6260
     |||   |||   ||||||:::   :::|||:::
1038 ArgSerProAlaAspValAlaAlaValLeuAsn..................  1048

6261 AACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAG 6310
           |||   |||         |||   :::      :::   :
1049 .......LysTyrTyrThrGlyAspValArgThr......HisAsnAlaT 1061

6311 TAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCT 6360
     :::::::::|||::::::   |||   |||   :::   ::::::|||:::
1061 hrAlaAsnSerMetThrValArgLysIleValSerLysGluGlnValSer 1077

6361 GAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
            :::      |||:::   :::   |||   ||||||:::::
1078 LeuLysProGlyAlaGlnTyrIleThrPheLeuGlnSerGluLysLysGl 1094

6411 TATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAG 6460
     ::::   :::   |||      :::         :::   |||   ||||
1094 uLeuValAsnLeuLeuAlaLeuArgLysValAlaAlaLysValSerThrV 1111

6461 TGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACG 6510
     ||||||||:::||||||:::|||||||:::||||||||||||   ||||||
1111 alHisGluSerGlnGlyGluThrPheLysAspValValLeuValArgThr 1127

6511 AAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGG 6560
     |||   :::|||||||||:::         :::   :::::::::|||::
1128 LysProThrAspAspSerIleAlaArgGlyArgGluTyrLeuIleValAl 1144

6561 TTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGT 6610
     :|||||||||||||||:::|||||||||||   :::::::::
1144 aLeuSerArgHisThrGlnSerLeuValTyrGluThrValLys....... 1158

6611 TGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCAGTA 6660
       :::|||   |||:::      |||   :::::::::      ::::::
1159 ..GluAspAspValSerLysGluIleArgGluSerAla......AlaLeu 1172

6661 TCCGACGCTTTGCTTCACACGTTC 6684
     :::   |||   |||      |||
1173 ThrLysAlaAlaLeuAlaArgPhe 1180
```

```
seq_name: pir1:WMFM12

seq_documentation_block:
125K protein - alfalfa mosaic virus
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase
C;Species: alfalfa mosaic virus, AMV
C;Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 05-May-2000
C;Accession: A04197
R;Cornelissen, B.J.C.; Brederode, F.T.; Moormann, R.J.M.; Bol, J.F.
Nucleic Acids Res. 11, 1253-1265, 1983
A;Title: Complete nucleotide sequence of alfalfa mosaic virus RNA 1.
A;Reference number: A04197; MUID:83143345
A;Accession: A04197
A;Molecule type: mRNA
A;Residues: 1-1126 <COR>
A;Cross-references: GB:L00163; GB:J02000; NID:g331424; PIDN:AAA46289.1; PID:g331425
C;Genetics:
A;Map position: segment 1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine
F;99-188/Domain: methyltransferase #status predicted <MTF>
F;838-845/Region: nucleotide-binding motif A (P-loop)
F;845-1096/Domain: helicase #status predicted <HHG>


alignment_scores:
            Quality:  329.50              Length:   1884
              Ratio:  0.504                 Gaps:     63
 Percent Similarity:  34.713   Percent Identity: 16.879

alignment_block:
 US-09-301-906-3 x WMFM12   ..

 Align seg 1/1  to: WMFM12  from: 1  to: 1126

    1303 ACGAAATTGTGCGATATGTTTTCTCAA..................AGGGA 1334
         |||   :::       :::|||||||:::                 :::||
      40 ThrAlaIleGlyLysValPheSerGluAlaGlyArgAlaTyrAlaGlnAs 56

    1335 CGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTT....... 1377
         ||||:::             ||||||   :::   :::|||:::
      56 pAlaLeu............ProSerAspLysGlyGluValLeuLysIleS 69

    1378 ........CTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCA 1419
                 |||         :::::::::   ||||||      ||||||
      69 erPheSerLeuAspAlaThrGlnGlnAsnIleLeuArgAlaAsnPhePro 85

    1420 GAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGC 1469
                       ||||||:::|||      |||||||||   |||||
      86 GlyArgArgThrValPheSerAsnSerSerSerSerSerHisCysPheAl 102

    1470 TAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTA 1519
         |   |||   |||         :::   :::      |||      ||||
     102 aAlaAlaHisArgLeuLeuGluThrAspPheValTyrArgCysPheGlyA 119

    1520 ATGTG......TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTC 1563
         ||:::         ::::::|||:::||||||:::|||:::   |||:::
     119 snThrValAspSerIleIleAspLeuGlyGlyAsnPheValSerHisMet 135

    1564 AAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAA 1613
         |||:::   :::   |||   |||   |||   |||:::||||||:::::
     136 LysValLysArgHisAsnValHisCysCysCysProIleLeuAspAlaAr 152

    1614 AGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGG 1663
         :|||      |||      ::::::||||||   |||:::
     152 gAspGlyAlaArgLeuThrGluArgIleLeuSerLeuLys..........  165

    1664 GAGATTCGTACGTGTCCAGT......GACCTTCTAACTGAAGCGGCTTCA 1707
              |||||||||   :::      :::::::::   ||||||
     166 .....SerTyrValArgLysHisProGluIleValGlyGluAla......  178

    1708 AAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGC 1757
                 :::|||||||            |||:::|||:::   |||||
```

```
 179 .........AspTyrCysMetAspThrPheGlnLysCysSerArgArgAl 192

1758 CGATGCGGGTTTTATGGTG...GATGTGTACGATATATCCCCGCAGCAGG 1804
     ||||   :::|||   :::   ::::::   |||::::::      ::::
 192 aAspTyrAlaPheAlaIleHisSerThrSerAspLeuAspValGlyGluL 209

1805 TAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTT 1854
     ::|||   ::::::|||:::|||||||::::::   |||      ::::::
 209 euAlaCysSerLeuAspGlnLysGlyValMetLysPheIleCysThrMet 225

1855 ATGTTCCCCGTGGAGTTGTTGTACGGTAAC...GGTGAAGTTTACTTGGA 1901
     |||      ::::::::::|||      |||   ||||||:::
 226 MetValAspAlaAspMetLeuIleHisAsnGluGlyGluIleProAsnPh 242

1902 AGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTAC.... 1947
      ::::::        :::   |||:::   |||   :::   :::
 242 eAsnValArgTrpGluIleAspArgLysLysAspLeuIleHisPheAspP 259

1948 ..............AATGTTGGTCAGTGTGGTGAGATGTATGAACATTCC 1983
                   |||:::|||               |||:::|||
 259 heIleAspGluProAsnLeuGly..............TyrSerHisArg 270

1984 TTCTCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGCACTTCGTC 2033
     ||||||   ::::::   ::::::|||:::
 271 PheSerLeuLeuLysHisTyrLeuThrTyr..................... 280

2034 CGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATC 2083
         |||:::      |||   :::   :::||||||
 281 ....AsnAlaValAspLeuGlyHisAlaAlaTyrArg............. 291

2084 ATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACT.........GAG 2124
             :::   |||   |||         |||:::        :::
 292 ........IleGluArgLysGlnAspPheGlyGlyValMetValIleAsp 305

2125 GTTACGTATAGGTCGTTG..............GTCCCGTCGTTCGTGGG 2159
     :::||||||   ||||||               :::|||         ||
 306 LeuThrTyr...SerLeuGlyPheValProLysMetProHisSerAsnGl 321

2160 CAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG....... 2202
     |:::|||   ::::::         |||   |||:::   |||
 321 yArgSerCysAlaTrpTyrAsnArgValLysGlyGlnMetValValHisT 338

2203 .........................TCCTTTAAGACAATAGTCCTCGAT 2226
                              |||::::::|||   |||
 338 hrValAsnGluGlyTyrTyrHisHisSerTyrGlnThrAlaValArgArg 354

2227 TCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGAC 2276
     :::   :::||||||:::               ::::::   ::::::
 355 LysValLeuValAspLysLysValLeuThrArgValThrGluValAlaPh 371

2277 ATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAA 2326
        :::   |||           |||   :::|||::::::|||
 371 eArgGlnPheArgProAsnAlaAspAlaHisSerAlaIleGlnSerIleA 388

2327 AGACC.....................CATGTCATTACAGGGAGTCGCGTT 2355
       |||                    |||:::|||   |||:::   :::
 388 laThrMetLeuSerSerSerThrAsnHisThrIleIleGlyGlyValThr 404

2356 GTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGT 2405
     :::   |||         :::||||||||||||         ::::::::
 405 LeuIleSerGlyLysProLeuSerProAspAspTyrIleProValAlaTh 421

2406 CGCTGTTATGGCTCAGGCGATTAAG........................ 2430
     :::::::      ::::::   |||
 421 rThrIleTyrTyrArgValLysLysLeuTyrAsnAlaIleProGluMetL 438

2431 ...........GATAGGGCGAAGAGTATTCGCTCCTATAACTTTATAAAA 2469
                |||:::::::::   :::   :::         :::|||
 438 euSerLeuLeuAspLysGlyGluArgLeuSerThrAspAlaValLeuLys 454

2470 GCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGACCGT 2519
     :::||||||||||   :::
 455 GlySerGluGlyProMet.....................TrpTyrSe 463

2520 AGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGC 2569
      |||      |||   :::|||::::::          |||:::   |||
 463 rGlyProThrPheLeuSerAlaLeuAsp.........LysValAsnValP 477

2570 ACGATAACTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTC 2619
          :::|||   |||   ::::::|||:::|||:::|||:::
 477 roGlyAspPhe...ValAlaLysAlaLeuLeuSerLeuProLys..... 490

2620 ATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGC 2669

 490 ................................................. 490

2670 TTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCT 2719

 490 ................................................. 490

2720 TCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATC 2769
          |||   ||||||:::             :::||||||||||:::
 491 .....ArgAspLeuLysSer............LeuPheSerArgSer... 500

2770 GTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAA 2819
                                                |||:::   :::::
 501 ................................AlaThrSerHisSerGl 506

2820 ACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACG 2869
     :      |||:::   ::::::
 506 uArgThrProValArgAspGlu............................ 513

2870 TTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCG 2919
       |||   ::::::|||
 514 ..SerProIleArgCys................................ 518

2920 ACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGAC 2969

 518 ................................................. 518

2970 GAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCT 3019
                                           |||   ::::::||||
 519 ................................ThrAspGlyValPheT 524

3020 ACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACA 3069
     ||
 524 yr.............................................. 524

3070 GAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGC 3119
                         |||:::   :::|||
 525 ..................ProIleArgMetLeu................. 529

3120 TGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCT 3169
           |||   :::::::::   |||       :::::::::|||::::
 530 ....LeuLysCysLeuGlySerAspLys......PheGluSerValThrI 543

3170 TAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATA 3219
     :::::|||
 543 leThrAsp.......................................... 545

3220 TTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGG 3269

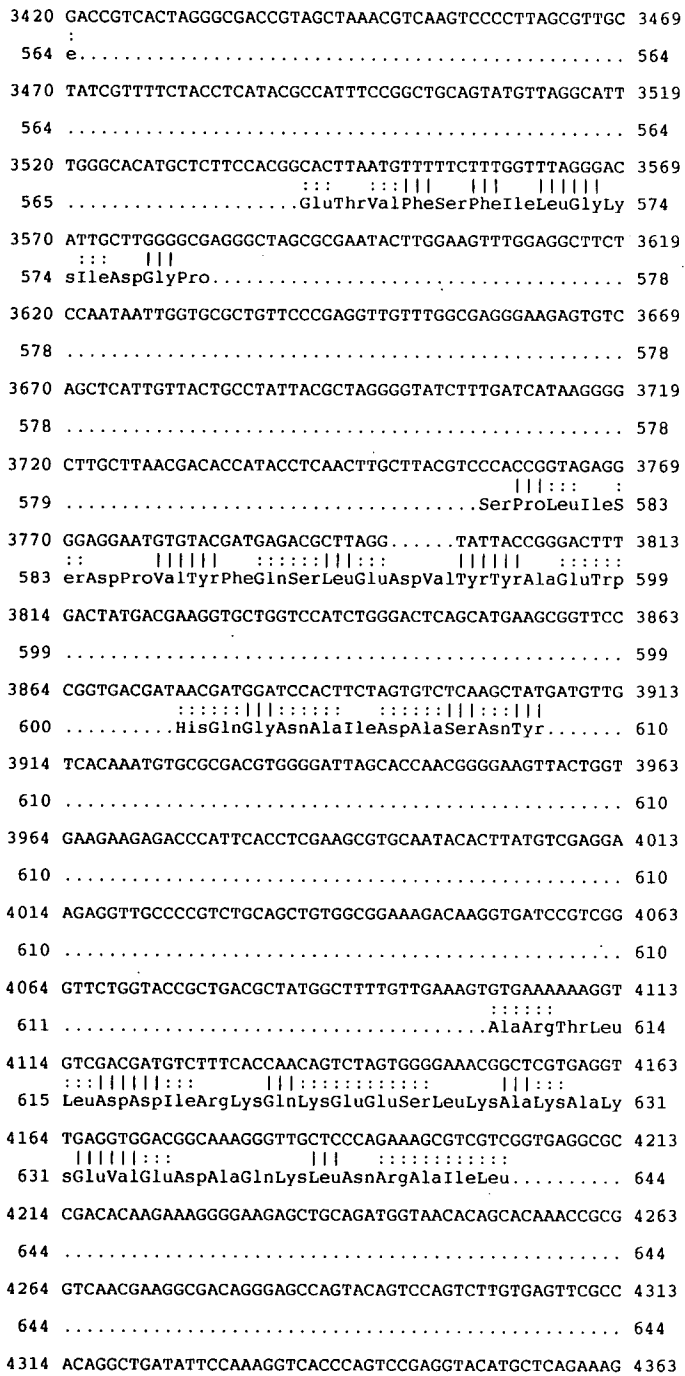
 545 ................................................. 545

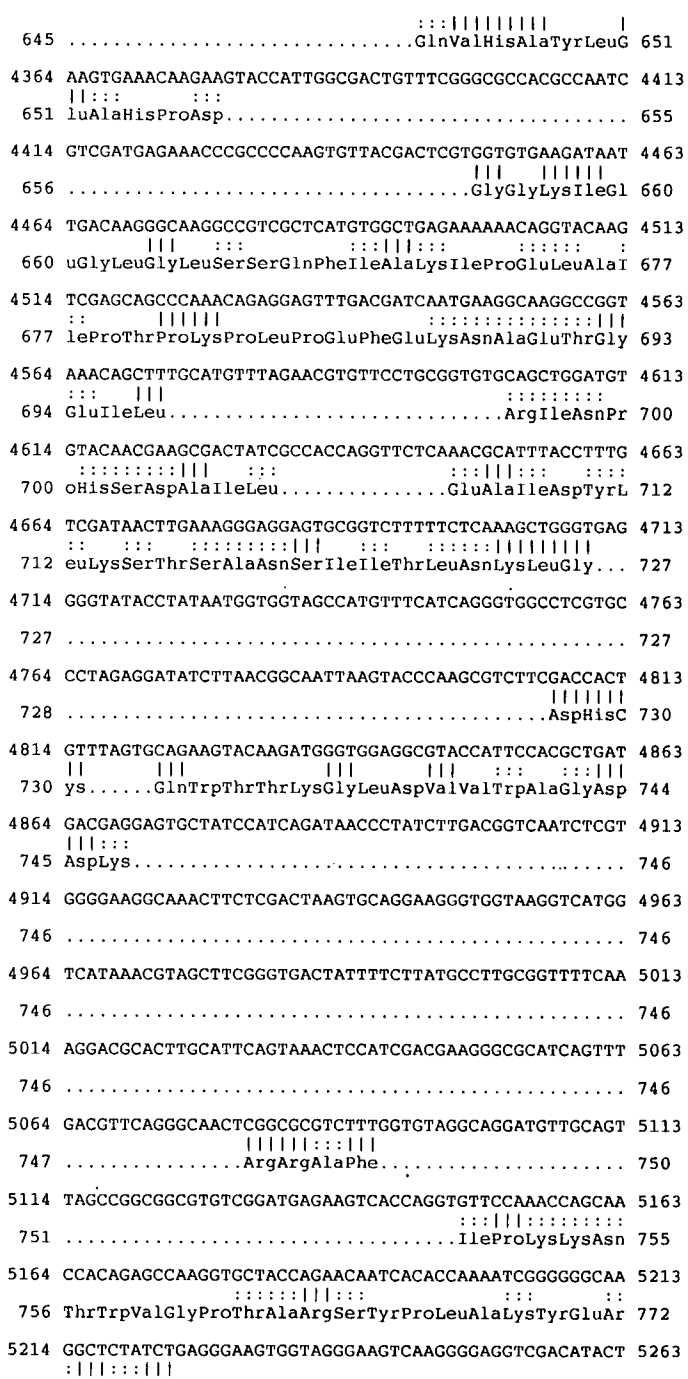
3270 AATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAA 3319

 545 ................................................. 545

3320 GCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGC 3369
                                        |||:::   |||::::::
 546 ............................ProArgSerAsnThrGluThr 552

3370 ACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGT 3419
     |||||||:::               :::   ::::::|||::::::   ::
 553 ThrValAsp..............LeuTyrGlnSerPheGlnLysLysIl 564
```


3420 GACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGC 3469
564 e 564
3470 TATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATT 3519
564 564
3520 TGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGAC 3569
565 GluThrValPheSerPheIleLeuGlyLy 574
3570 ATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCT 3619
574 sIleAspGlyPro 578
3620 CCAATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTC 3669
578 578
3670 AGCTCATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGG 3719
578 578
3720 CTTGCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGG 3769
579 SerProLeuIleS 583
3770 GGAGGAATGTGTACGATGAGACGCTTAGG TATTACCGGGACTTT 3813
583 erAspProValTyrPheGlnSerLeuGluAspValTyrTyrAlaGluTrp 599
3814 GACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCC 3863
599 599
3864 CGGTGACGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTG 3913
600 HisGlnGlyAsnAlaIleAspAlaSerAsnTyr 610
3914 TCACAAATGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGT 3963
610 610
3964 GAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGA 4013
610 610
4014 AGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGG 4063
610 610
4064 GTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGT 4113
611 AlaArgThrLeu 614
4114 GTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGT 4163
615 LeuAspAspIleArgLysGlnLysGluGluSerLeuLysAlaLysAlaLy 631
4164 TGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGC 4213
631 sGluValGluAspAlaGlnLysLeuAsnArgAlaIleLeu 644
4214 CGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCG 4263
644 644
4264 GTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCC 4313
644 644
4314 ACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAG 4363


645 GlnValHisAlaTyrLeuG 651
4364 AAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATC 4413
651 luAlaHisProAsp 655
4414 GTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAAT 4463
656 GlyGlyLysIleGl 660
4464 TGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAG 4513
660 uGlyLeuGlyLeuSerSerGlnPheIleAlaLysIleProGluLeuAlaI 677
4514 TCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGT 4563
677 leProThrProLysProLeuProGluPheGluLysAsnAlaGluThrGly 693
4564 AAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGT 4613
694 GluIleLeu ArgIleAsnPr 700
4614 GTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTG 4663
700 oHisSerAspAlaIleLeu GluAlaIleAspTyrL 712
4664 TCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAG 4713
712 euLysSerThrSerAlaAsnSerIleIleThrLeuAsnLysLeuGly... 727
4714 GGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGC 4763
727 727
4764 CCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACT 4813
728 AspHisC 730
4814 GTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGAT 4863
730 ys GlnTrpThrThrLysGlyLeuAspValValTrpAlaGlyAsp 744
4864 GACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGT 4913
745 AspLys 746
4914 GGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGG 4963
746 746
4964 TCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAA 5013
746 746
5014 AGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTT 5063
746 746
5064 GACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGT 5113
747 ArgArgAlaPhe 750
5114 TAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAA 5163
751 IleProLysLysAsn 755
5164 CCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAA 5213
756 ThrTrpValGlyProThrAlaArgSerTyrProLeuAlaLysTyrGluAr 772
5214 GGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACT 5263

```
 772 gAlaMetSer..................................... 775

5264 CGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313
               ::::::   ||||||              ||||||
 776 ...........LysAspGlyTyrValThr............LeuArgTrp 784

5314 GATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGA 5363
     |||:::   |||:::      :::         :::         :::||
 785 AspGlyGluValLeuAspAlaAsnCysValArgSerLeuSerGlnTyrGl 801

5364 AGTGGTTAAAACCGGG..............ACCTCTGAAGATGCCGTCG 5398
     |:::|||   :::                  :::|||   :::|||:::
 801 uIleValPheValAspGlnSerCysValPheAlaSerAlaGluAlaIleI 818

5399 TGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCG 5448
     ::      |||:::                          :::|||
 818 leProSerLeuGlu............................LysAla 824

5449 TTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTG.. 5496
     |||   :::      :::::::::||||||   ::::::||||||
 825 LeuGlyLeuGluAlaHisPheSerValThrIleValAspGlyValAlaGl 841

5497 .......AAAGTACCTAAT......CAAGTTTATGAATCACTACCGGGCT 5533
            |||:::   |||      |||:::   :::|||
 841 yCysGlyLysThrThrAsnIleLysGlnIleAlaArgSer.......... 854

5534 TTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGAC 5583

 854 ................................................. 854

5584 GGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTAT 5633

 854 ................................................. 854

5634 CTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATC 5683
            :::|||:::|||||||||   :::      |||   :::|||
 855 .......SerGlyArgAspValAspLeuIle......LeuThrSerAsnA 867

5684 TGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATA 5733
            :::|||::::::          |||:::      |||
 867 rgSerSerAlaAspGluLeuLysGluThrIleAspCysSerProLeuThr 883

5734 GGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTC 5783
                :::      |||
 884 LysLeuHisTyrIleArgThrCys........................... 891

5784 CTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGA 5833

 891 ................................................. 891

5834 CGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACG 5883
             |||::::::
 892 ........AspSerTyr.............................. 894

5884 GCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGA 5933

 894 ................................................. 894

5934 TCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGG 5983
                   |||   ||||||::::::|||::::::
 895 ...............LeuMetSerAlaSerAlaValLysAla....... 903

5984 TAAACTTTATCGTCAGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTG 6033
                            :::|||::::::   ||||||
 904 .........................GlnArgLeuIlePheAspGluCys 911

6034 CACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGC 6083
        ::::::|||   ||||||::      ::::::      :::|||::
 912 PheLeuGlnHisAlaGlyLeuValTyrAlaAlaAlaThrLeuAlaGlyCy 928

6084 GTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACA 6133
     :||||||   :::   |||||||||   :::|||||||||||::::::|
 928 sSerGluValIleGlyPheGlyAspThrGluGlnIleProPheValSerA 945

     6134 GGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAA 6183
          ||:::      |||   :::            ::::::   |||   |||
      945 rgAsnProSerPheValPheArgHisHisLysLeuThrGlyLysValGlu 961

     6184 AGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTA 6233
               :::   |||         :::|||  |||   |||:::   ||
      962 ArgLysLeuIleThr.........TrpArgSerProAlaAspAlaThrTy 975

     6234 CTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAA 6283
          |   |||                                      ||||
      975 rCysLeu.....................................GluL 979

     6284 AGGTCGTTAGCGGTAAGGACAAACCA..................GTAGTA 6315
          ||            :::   ||||||                 |||:::
      979 ysTyrPheTyrLysAsnLysLysProValLysThrAsnSerArgValLeu 995

     6316 AGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAAT 6365
          ||||||::::::      ||||||::::::      :::|||   :::
      996 ArgSerIleGluValValProIleAsnSerProValSerValGluArgAs 1012

     6366 AAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGA 6415
           :::      :::|||||||||   ||||||   ||||||:::   :::|
     1012 nThrAsnAlaLeuTyrLeuCysHisThrGlnAlaGluLysAlaValLeuL 1029

     6416 AGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCAT 6465
          ||   :::         ||||||   ::::::   ::::::|||:::|||
     1029 ysAlaGlnThrHisLeuLysGlyCysAspAsn...IlePheThrThrHis 1044

     6466 GAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAA 6515
          ||||||||||||||||||||::::::|||   :::   |||      ::
     1045 GluAlaGlnGlyLysThrPheAspAsnValTyrPheCysArgLeuThrAr 1061

     6516 AGCCGATGACTCCCTATTCACTAAACAACCGCATATA............. 6552
          :::::::   ||||||   |||   :::      |||
     1061 gThrSerThrSerLeuAlaThrGlyArgAspProIleAsnGlyProCysA 1078

     6553 .....CTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCT 6597
               ||||||:::||||||||||||   :::::::::   |||   :::
     1078 snGlyLeuValAlaLeuSerArgHisLysLysThrPheLysTyrPheThr 1094

     6598 CTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTC 6647
          ::::::         ||||||   :::   :::         ||||||::
     1095 IleAlaHisAspSerAspAspValIleTyrAsnAlaCysArgAspAlaGl 1111

     6648 ACCTCAATCAGTATCCGACGCTTTGCTT..............CACACGT 6682
          :      :::::::::|||::::::|||               |||:::|
     1111 y......AsnThrAspAspSerIleLeuAlaArgSerTyrAsnHisAsnP 1126

     6683 TC 6684
          ||
     1126 he 1126
```

```
seq_name: pir2:T45025

seq_documentation_block:
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A;Reference number: Z22899; MUID:97166151
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B
```

```
alignment_scores:
           Quality:  320.50              Length:   2469
             Ratio:  0.309                 Gaps:    118
 Percent Similarity:  41.960   Percent Identity: 20.697

alignment_block:
 US-09-301-906-3/rev x T45025   ..

 Align seg 1/1  to: T45025  from: 1  to: 3570

    6430 CCTTCAACGACCTCTTCA......TATCCGACTTCTCCAACTGGGTCATG 6387
         |||:::::|||||||||         |||:::|||      |||:::::
    1150 ProThrValThrSerSerLysAlaThrProSerSerSerProGlyThrAl 1166

    6386 CACAAGTACACGTCAGCGTTTATTTCAGCTACGTCATCAGTGGTTCCAAT 6337
         :|||:::           |||      ||||||
    1166 aThrAla............LeuProAlaLeuArgSer............. 1174

    6336 TGGCCTTTTGGACAGCGATCTTACTACTGGTTTGTCCTTACCGCTAACGA 6287
                  ||||||
    1175 ...........ThrAlaThr.........................Thr 1178

    6286 CCTTTTCAGGGTAACACTTTTCCGTTCCCCTTACGGTCATTGAGGACAAC 6237
         |||         ::::::   |||:::|||      ||||||   |||||
    1179 ProThrAlaThrSerValThrProIleProSerSerSerLeuGlyThrTh 1195

    6236 AAGTAGCAAACATCTAACGGACACCTGTACGATTTAGAAGTGTATACAAC 6187
         |   ::::::   |||:::::|||   |||
    1195 rTrpThrArg...LeuSerGlnThrThrThr.................. 1204

    6186 GCTTTCCTTCTTTGGAACAAAAACAGCACAATCCATCCTAAACACCT... 6140
              |||:::      ::::::      :::||||||   ||||||
    1205 .....ProThrAlaThrMetSerThrAlaThrProSerSerThrProGlu 1219

    6139 ............TCTCCCTGTTTATGAATGGTATCTGA............ 6114
                     |||      |||      ::::::
    1220 ThrAlaHisThrSerThrValLeuThrAlaThrAlaThrThrThrGlyAl 1236

    6113 ..........TTTATGTCTCCAAAAAAGAGGCCTTCCGACGCGCCGGTTG 6074
                        :::|||::::::   |||:::   |||.  ::::
    1236 aThrGlySerValAlaThrProSerSerThrProGlyThrAlaHisThrT 1253

    6073 CGAAGACGCCTAGT..................TGTAGTAAGCCTTGATGC 6042
         ::|||:::|||:::                    :::   |||
    1253 hrLysValProThrThrThrThrThrGlyPheThrAlaThrPro...... 1267

    6041 ATCATGTGCACCTCATCCACCAAAACCCTTTTATACATTCCCCTGACGAT 5992
                  :::|||||||      |||   |||      |||:::   ||
    1268 .........SerSerSerProGlyThrAlaLeuThrProProValTrpIl 1281

    5991 AAAGTTTACCACCCTGGAGTTAACTGTGGTAGCACTGTTGAGACCTTCCA 5942
         |:::   ||||||         |||      :::   :::   ||||||:
    1281 eSerThrThrThrThrProThrThrArgGlySerThrValThrProSerS 1298

    5941 AATTCGGATCTCTCTTCTTCACCGCCATATTTATGTCCTCAGAACTTCCC 5892
         :::::   :::         ||||||      :::::::::
    1298 erIleProGlyThrThrHisThrAlaThrValLeuThrThrThrThrThr 1314

    5891 ACATTAGCCGTAATGGTGGCTGTGCTATTGGGTGACTTAACGAATTCGTC 5842
         |||:::|||:::                  |||:::::::::::   ||
    1315 ThrValAlaThr.................GlySerMetAlaThrProSe 1325

    5841 CACTAACGTCGTCGTCTTACCGCCACCTGGGGGAGCATTATAGCACTTAT 5792
         |:::::::::   :::          |||   :::
    1325 rSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaT 1342

    5791 ATTCGAAGGACTTAAACGATTCGCCCACCATGCCGCATCGAGCGACTTTC 5742
           :::         :::   |||   |||   |||      :::
    1342 hrThrIleThrAlaThrGlySerThrThrAsnProSerSerThrProGly 1358

    5741 AGTGCACCTATC...AAATTAATGGCATCGTGGAAAACCAGTATATCGTC 5695
         :::   ||||||      ::::::::::::      |||:::   :::::

    1359 ThrArgProIleProProValLeuThrThrThrAlaThrThrProAlaAl 1375

    5694 GCATACGAACAGATTGTCGCCCAAAGCTACCACCGCGTCGACATCTTTGC 5645
         :   ::::::   ::::::|||:::::::::   ::::::|||      |
    1375 aThrSerSerThrValThrProSerSerAlaLeuGlyThrThrHisThrP 1392

    5644 CCTTGGTAAAGATACCTTTTTCAGCTATGAGTACGTACGGTAGGTCTCTC 5595
         ||         :::|||   ::::::   :::|||:::||||||||||||
    1392 ro......ProValProAsnThrThrAlaThrThrHisGlyArgSerLeu 1406

    5594 ACACGCAAGCCGTCTTGTGTTGAATGAAAAATGAGATCTGTGCCCGACTT 5545
         :::      ||||||
    1407 Ser......ProSer..................................... 1409

    5544 GTAAACGTGAAAGCCCGGTAGTGATTCATAAACTTGATTAGGTACTTTCA 5495

    1409 ..................................................... 1409

    5494 GAACACCTTCGGCGGTAGTGACGGCAATATTTCTAGCCATAAGCAACGCC 5445
           :::|||   :::|||   ||||||         |||   |||::::::
    1410 ..SerProHisThrValArgThrAlaTrpThrSerAlaThrSerGlyThr 1425

    5444 CTGTATGTCCTCTCAATGCCTATAGCCAGATACTTCAAGTACTCCACGAC 5395
         |||                      :::   ::::::         :::||
    1426 LeuGly.................ThrThrHisIleThrGluProSerTh 1436

    5394 GGCATCTTCAGAGGTCCCGGTTTTAACCACTTCAAATGTCATTGGGGTGT 5345
         |:::::::|||::::::|||:::   ||||||::::::::
    1436 rGlyThrSerHisThrProAlaAlaThrThrGlyThrThr.......... 1449

    5344 AGTCAGGTTCAAGAGCCATCACTGGATTATCAGCCCTGAGCCACTCACAC 5295
                                                :::||||||
    1450 ....................................GlnHisSer... 1452

    5294 TTCCTAACGTAATCTTGTTCGCACCATATCGAGTATGTCGACCTCCCCTT 5245
                                                      |||
    1453 ..........................................ThrProAl 1455

    5244 GACTTCCCTACCACTTCCCTCAGATAGAGCCTTG......CCCCCCGATT 5201
         |||   |||   ||||||:::|||:::         ||||||:::
    1455 aLeuSerSerProHisProSerSerArgThrThrGluSerProProSerP 1472

    5200 TTGGTGTGATTGTTCTGGTAGCACCTTGGCTCTGTGGTTGCTGGTTTGGA 5151
           |||:::
    1472 roGlyThrThr..................................... 1475

    5150 ACACCTGGTGACTTCTCATCCGACACGCCGCCGGCTAACTGCAACATCCT 5101
         |||||||||               |||      :::::   :::
    1476 ThrProGly.........HisThrThrAlaThrSerArgThrThrAlaTh 1489

    5100 GCCTACACCA...AAGACGCGCCGAGTTGCCCTGAACGTCAAACTGATGC 5054
         |||||   |||||||||                 ::::::   :::
    1489 rAlaThrProSerLysThrArg..............ThrSerThrLeuL 1501

    5053 GCCCTTCGTCGATGGAGTTTACTGAATGCAAGTGCGTCCTTTGAAAACCG 5004
           |||||||||                                  |||
    1501 euProSerSer.................................Pro 1505

    5003 CAAGGCATAAGAAAATAGTCACCCGAAGCTACGTTTATGACCATGACCTT 4954
           :::            :::|||   :::|||   :::||||||
    1506 ThrSer............AlaProIleThrThrValValThrMetGlyCy 1518

    4953 ACCACCCTTCCTGCACTTAGTCGAGA............AGTTTGCCTTCC 4916
           :::|||:::|||:::   |||:::                |||   |||
    1518 s.GluProGlnCysAlaTrpSerGluTrpLeuAspTyrSerTyrProMet 1534

    4915 CCACGAGATTGACCGTCAAGATAGGGTTATCTGATGGATAGCACTCCTCG 4866
         |||         ||||||         :::         |||:::   ||
    1535 ProGly......ProSerGlyGlyAspPheAspThrTyrSerAsnIleAr 1549

    4865 TCATCAGCGTGGAATGGTACGCCTCCACCCATCTTGTACTTCTGCACTAA 4816
         |         |||   |||      :::|||
    1549 gAlaAlaGlyGlyAlaValCysGluGlnPro.................... 1559
```

```
4815 ACAGTGGTCGAAGACGCTTGGGTACTTAATTGCCGTTAAGATATCCTCTA 4766
                     |||||| |||:::||||||
1560 ................LeuGly.LeuGluCysArg..............  1565

4765 GGGCACGAGGCCACCCTGATGAAACATGGCTACCACCATTATAGGTATAC 4716
                                |||   :::      |||:::|
1566 ............................AlaGlnAlaGlnProGlyValP 1573

4715 CCCTCACCCAGCTTTGAGAAAAAGACCGCACTCCTCCCTTTCAAGTTATC 4666
     |||||   :::|||                              ||||||:::
1573 roLeuArgGluLeu.......................GlyGlnValVal 1581

4665 GACAAAGGTAAATGCGTTTGAGAACCTGGTGGCGATAGTCGCTTCGTTGT 4616
     :::         |||      :::   |||    :::
1582 Glu.........CysSerLeuAspPheGly.LeuVal.............  1590

4615 ACACATCCAGCTGCACACCGCAGGAACACGTTCTAAACATGCAAAGCTGT 4566
                |||      :::|||:::|||   :::::::::   |||
1591 ...........CysArgAsnArgGluGlnValGlyLysPheLysMetCys 1603

4565 TTACCGGCCTTGCCTTCATTGATCGTCAAACTCCTCTGTTTGGGCTGCTC 4516
     :::            :::      ::::::::::::|||   :::   ::
1604 Phe...........AsnTyrGluIleArgValPheCysCysAsnTyrGl 1616

4515 GACTTGTACCTGTTTTTTCTCAGCCACATGAGCGACGGCCTTGCCCTTGT 4466
     :   |||               ||||||   :::||||||   |||   |
1616 yHisCysProSer...ThrProAlaThrSerSerThrAlaThrProSerS 1632

4465 CAATTATCTTCACACCACGAGTCGTAACACTTGGGGCGGGTTTCTCATCG 4416
     ||        |||      ::::::|||
1632 erThrProGlyThrThrTrpIleLeuThr..................... 1641

4415 ACGATTGGCGTGGCGCCCGAAACAGTCGCCAATGGTACTTCTTGTTTCAC 4366
                    :::|||:::|||      |||::::::   ||
1642 ..............GluGlnThrThrAlaAlaThrThrThrAlaThrTh 1653

4365 TTCTTTCTGAGCATGTACCTCGGACTGGGTGACCTTTGGAATATCAGCCT 4316
     |:::      |||      |||:::      |||   |||   :::
1653 rGlySerThrAlaIleProSerSer......ThrProGlyThrAlaProP 1668

4315 GTGGCGAACTCACAAGACTGGACTGTACTGGCTCCCTGTCGCCTTCGTTG 4266
          ::::::      ::::::|||          :::   |||
1668 roProLysValLeuThrSerGlnAlaThrThrProThrAlaThrSerSer 1684

4265 ACCGCGGTTTGTGCTGTGTTACCATCTGCAGCT......CTTCCCCTTTC 4222
        |||:::   :::      |||   :::|||      |||||||:::
1685 LysAlaThrSerSerSerSerProArgThrAlaThrThrLeuProValLe 1701

4221 TTGTGTCGGCGCCTCACCGACGACGCTTTCTGGGAGCAACCCTTTGCCGT 4172
               |||:::   :::|||   ::::::   :::|||:::||||
1701 uThrSerThrAlaThrLysSerThrAlaThrSerPheThrProIleProS 1718

4171 CC....................ACCTCAACCTCACGAGCCGTTTCCCCA 4143
     ||                    ||||||   :::|||         |||
1718 erSerThrLeuGlyThrThrGlyThrSerGlnAsnArgProProHisPro 1734

4142 CTAGACTGTTGG......TGAAAGACATCGTCGACACCTTTTTTCACA.. 4101
     :::                          ||||||||||||      |||
1735 MetAlaThrMetSerThrIleHisProSerSerThrProGluThrThrHi 1751

4100 .............CTTTCAACAAAAGCC............ATAGCGTCAG 4076
                  |||:::|||||||||               :::|||:
1751 sThrSerThrValLeuThrThrLysAlaThrThrThrArgAlaThrSerS 1768

4075 CGGTACCAGAACCCGACGGATCACCT....................... 4050
     :::::      |||::::::::::|||
1768 erMetSerThrProSerSerThrProGlyThrThrTrpIleLeuThrGlu 1784

4049 .........TGTCTTTCCGCCACAGCTGCA......GACGGGGCAACCTC 4015
              :::   ::::::|||||||||         |||:::   ||
1785 LeuThrThrAlaAlaThrThrThrAlaAlaLeuProHisGlyThrProSe 1801
```

```
4014 TTCCTCGACATAAGTGTATTGCACGCTTCGAGGTGAATGGGTCTCTTCTT 3965
     ||||                  |||           |||:::   ::::
1801 rSer.................ThrProGlyThrThrTrpIleLeuThrG 1812

3964 CACCAGTAACTTCCCCGTTGGTGCTAATCCCCACG.............. 3930
     ::|||   |||:::      |||   :::||||||
1812 luProSerThrThrAlaThrValThrValProThrGlySerThrAlaThr 1828

3929 ...TCGCGCACATTTGTGACAACATCATAGCTTGAGACACTA........ 3891
        |||   |||   :::|||::::::   |||::::::|||
1829 AlaSerSerThrArgAlaThrAlaGlyThrLeuLysValLeuThrSerTh 1845

3890 ............................GAAGTGGATCCATCGTTATCGT 3869
                                 ::::::   ||||||   ||||
1845 rAlaThrThrProThrValIleSerSerArgAlaThrProSer...SerS 1861

3868 CACCGGGAACCGCTTCATGCTGAGTCCCAGATGGACCAGCACCTTCGTCA 3819
     ||||||||||||||::::::                |||||      ||
1861 erProGlyThrAlaThrAla............LeuProAlaLeuArgSer 1873

3818 TAGTCAAAGTCCCGGTAATACCTAAGCGTCTCATCGTACACATTCCTCCC 3769
        :::   :::              :::::::::|||   |||   :::||
1874 ThrAlaThrThrPro.........ThrAlaThrSerValThrAlaIlePr 1887

3768 CTCTACC.........................GGTGGGACGTAAGCAA 3746
     ||||:::                       :::   |||   :::
1887 oSerSerSerLeuGlyThrAlaTrpThrArgLeuSerGlnThrThrThrP 1904

3745 GTTGAGGTATGGTGTCGTTAAGCAAGCCCCTTATGATCAAAGATACCCCT 3696
          :::   :::|||   :::   |||          ::::::||||||
1904 roThrAlaThrMetSerThrAlaThrPro.........SerSerThrPro 1917

3695 AGCGTAATAGGCAGTAACAATGAGCTGACACTCTTCCCTCGCCAAACA.. 3648
     :::::::::   :::::::::   ||||||               |||
1918 GluThrValHisThrSerThrValLeuThrThrThrAlaThrThrThrAr 1934

3647 ............................ACCTCGGGAACAGCGCACCAAT 3626
                                  |||   ||||||||||||
1934 gThrGlySerValAlaThrProSerSerThrProGlyThrAlaHis.... 1949

3625 TATTGGAGAAGCCTCCAAACTTCCAAGTATTCGCGCTAGCCCTCGCCCCA 3576
           |||   |||:::      :::   :::::::::   :::|||
1950 ..ThrThrLysValProThrThrThrThrThrGlyPheThrAlaThrPro 1965

3575 AGCAATGTCCCTAAACCAAAGAAAAACATTAAGTGCCGTGGAAGAGCATG 3526
     |||:::   |||                          |||   |||
1966 SerSerSerPro.........................GlyThrAlaLe 1973

3525 TGCCCAA.........ATGCCTAACATACTGCAGCCGGAAATGGCGTATG 3485
      :::            :::   :::          |||           :
1973 uThrProProValTrpIleSerThrThrThrThrProThrThrArgGlyS 1990

3484 AGGTAGAAAACGATAGCAACGCTAAGGGGACTTGACGTTTAGCTACGGTC 3435
     :::::   :::   |||:::      ||||||   :::   |||||||||
1990 erThrValThrProSerSerIleProGlyThrThrHisThrAlaThrVal 2006

3434 GCCCTAGTGACGGTCACC.....................GCCGACGA 3409
           :::|||:::|||                      :::   ::
2007 LeuThrThrThrThrThrThrValAlaThrGlySerMetAlaThrProSe 2023

3408 AAACCCTCTTCGCACGACTAAAGATCCACCGGAT.............. 3375
     ::::      :::|||:::      ||||||:::
2023 rSerSerThrGlnThrSerGlyThrProProSerLeuThrThrThrAlaT 2040

3374 ............ACTGTGCTAGAGGTTAAACCGCGTAAACCACCGGCG 3339
                  |||         ::::::|||   :::   |||:::
2040 hrThrIleThrAlaThrGlySerThrThrAsnProSerSerThrProGly 2056

3338 GCTTGTGAACTTAACGAGCTTAGAGTTTCGTAACTACTGCCCGCGTACCT 3289
     :::                                :::|||      ||
2057 ThrThrPro.........................IleProProValLe 2064

3288 TCTAGCGTTAGCCGCAATTCCGCGCCTATACTTCCATGCAATTGGGGCCG 3239
```


2064 uThrThrThrAlaThrThrProAlaAlaThrSerSerThrVal.....T 2079
3238 CACCCAAAGACGCCACGAAT.....ATCCCGATAAGC 3207
2079 hrProSerSerAlaLeuGlyThrThrHisThrProProValProAsnThr 2095
3206 ACCTTATTCCGCAATACGAAAGTGGTGAAGTCATTTAAGGAAACTTCACT 3157
2096 ThrAlaThrThrHisGlyArgSerLeuProProSerSerProHisThrVa 2112
3156 CCAAAATTCTTTCGTTCCAGCTACGACCTTGCTCACAGCTACGCGCGACG 3107
2112 lProThrAlaTrpThrSerAlaThrSerGlyIleLeuGlyThrThrHisI 2129
3106 TCTCCAGTATGCCCACCGGTGAGTTTATAGCATTCTCTGTGCGCACCAGA 3057
2129 leThrGluProSerThrGlyThrSerHisThrProAlaAlaThrThrGly 2145
3056 GCATCAAGCCTTTCA.....TCTTCAGTTAAACCGCGGTCGTA 3019
2146 ThrThrGlnProSerThrProAlaLeuSerSerProHisProSerSerAr 2162
3018 GAATAGTGAA...CCTGTGGAAAAGAACGACGTAATGCCA..... 2982
2162 gThrThrGluSerProProSerProGlyThrThrThrProGlyHisThrA 2179
2981 GAAAAGGCCTTCGTCGCAGCCCCACGAGCTACCAAA... 2946
2179 rgGlyThrSerArgThrThrAlaThrAlaThrProSerLysThrArgThr 2195
2945 CCCTTAGAAACACTCGA 2929
2196 SerThrLeuLeuProSerSerProThrSerAlaProIleThrThrValVa 2212
2928 AACCACGGTCGCTTTCGAACCGCCAAGAAGACCTGCCCTCGTACAGTGGA 2879
2212 lThrThr...GlyCysGluPro.....GlnCysAlaTrpSerGluTrpL 2226
2878 CGTTAGAAACGTTCCCCATT.....ATGAACGCA 2850
2226 euAspTyrSerTyrProMetProGlyProSerGlyGlyAspPheAspThr 2242
2849 TATACAGAATCTTCAGTAATTGGC..... 2826
2243 TyrSerAsnIleArgAlaAlaGlyGlyAlaValCysGluGlnProLeuGl 2259
2825 AGGGGTTTTCCATCTTCTGTCTTCATTGCCTTCA 2792
2259 yLeuGluCysArgAlaGlnAlaGlnProGlyValProLeuArgGluLeuG 2276
2791 TGACCTTAATTGAATCTTCC...ACGATAGCCCTAGAGAAGACGCGCAAG 2745
2276 lyGlnValValGluCysSerLeuAspPheGlyLeuValCysArgAsnArg 2292
2744 CGACTATTCTTCAGTTTCCTACCGAAGGTCTCCTTCGAAATATCAAAGGC 2695
2293 GluGlnValGlyLysPhe.....LysMetCysPheAsnTyrGluIleAr 2307
2694 ACCCCTGAGCTCCAACCTGTCTGAAGCTCCAGAAGTGCAAATGGTAACAA 2645
2307 gValPheCysCysAsnTyrGlyHisCysProSerThrProAlaThrSerS 2324
2644 CAACA.....GACCCAGGTACTTTACGGATGAACGCTCTG 2610
2324 erThrAlaThrProSerSerThrProGlyThrThrTrp..... 2336
2609 GGCATAGACATAAGCGTCTCCAAAACGTTGAAGTTATCGTGCACCATTGC 2560
2337 IleLeuThrLysLeuThrThrThrAlaThrThrThrGluSe 2350
2559 CTTAGCATAGACGGAGACTGCGTTCGAAAAACAATCGCCTACGGTCTGAA 2510


2350 rThrGlySerThrAlaThrProSerSerThrGlnGlyProProAlaGlyT 2367
2509 AGAAGAGCTTGAAGACCCCGGCGAGACTCCCCTCACTGGCTTTTATAAAG 2460
2367 hrProHisValSerThrThrAlaThrThrProThrValThrSerSerLys 2383
2459 TTA.....TAGGAGCGAATACTCTTCGCCCTATC 2431
2384 AlaThrProPheSerSerProGlyThrAlaThrAlaLeuProAlaLeuAr 2400
2430 CTTAATCGCCTGAGCCATAACAGCGACAACTAAACCCCACATATCATCAG 2381
2400 gSerThrAlaThrThrProThrAlaThrSerPheThrAlaIleProSerS 2417
2380 GAGAAATATCAACCTTGCTGTGG.....ACAACGCGA 2349
2417 erSerLeuGlyThr...ThrTrpThrArgLeuSerGlnThrThrThrPro 2432
2348 CTCCCTGTAATGACATGGGTCTTTTGCGACCTGACCGCCCCAACGGCATA 2299
2433 MetAlaThrMetSerThrAlaThrProSerSerThrProGluThrValHi 2449
2298 CTCAAACGTTCTATTCTCGAATGTCCCTATAGTGTTGAGCGCGTAGGAAT 2249
2449 sThrSerThrValLeuThrThrThrAlaThrThrThrGlyAlaThrGlyS 2466
2248 AGATCCTGTCGACAAAGTCCGAATCGAGGACTATTGTCTTAAAGGACACA 2199
2466 erValAlaThrProSerSerThrProGlyThrAlaHisThrThrLysVal 2482
2198 CTAGAACCAGCTACAACAGGT..... 2178
2483 ProThrThrThrThrThrGlyPheThrValThrProSerSerSerProGl 2499
2177 ATGAACACCAGCGATTTGCCCACGA 2153
2499 yThrAlaArgThrProProValTrpIleSerThrThrThrThrProThrT 2516
2152 ACGACGGGACCAACGACCTATACGTAACCTCAGTTCCAGGTGACTTCTGA 2103
2516 hrSerGlySerThr...ValThrProSerSerIleProGly..... 2528
2102 GCCCTACACATAGTGAGATGATGGTAACCACAACGGTATCCCTCATACTC 2053
2529 Thr..ThrHisThrProThrValLeuThrThrThrThrGlnProValAla 2544
2052 TAGCTTAAACACGTTCCCGGACGAAGTGCGTACATAAGAAAAGGTGAAAA 2003
2545 ThrGlySerMetAlaThrProSerSerSerThrGlnThrSerGlyThrPr 2561
2002 ACCCGCTTACGTTAGAGAAGGAATGTTCATACATCTCACCACACTGACCA 1953
2561 oProSerLeuIle.....ThrThrAlaThrT 2570
1952 ACATTGTAGGCCAGGTAATCACCTTCCCTCTTCACCAACGTATCGAGTTC 1903
2570 hrIleThr.AlaThrGlySer.....ThrThrAsnProSerSe 2582
1902 TTCCAAGTAAACTTCACCGTTACCGTACAACAACTCCACGGGGAACATAA 1853
2582 rThrProGlyThrThrProIlePro..... 2590
1852 GAGCTATGTCGAAAACCAGCGCACCCTTCTTATCCATAGCCTCTGCTACC 1803
2591 ProGluLeuThrThrThrAlaThrThr 2599
1802 TGCTGCGGGGATATATCGTACACATCCACCATAAAACCCGCATCGGCTCT 1753
2600 ProAlaAlaThrSerSerThrValThrProSerSerAlaLe 2613
1752 AGAATCGCAGTTCTGCGATTCTCGACTACAGTAAGACACAGACTTTGAAG 1703
2613 uGlyThr.....ThrHisThrProProValProAsnThrThrA 2626

```
1702 CCGCTTCAGTTAGAAGGTCACTGGACACGTACGAATCTCCCCCAGCTGTG 1653
     ||:::::      |||||||||            ||||||   :::
2626 laThrThrHisGlyArgSerLeu.....SerProSerSerPro 2638

1652 GAAAGAAAGAGGATCTCATTGATTCTCCGCTTCACATCTTTAACGTCTAG 1603
     :::      |||   :::           :::||||||   ||||||
2639 HisThrValArgThrAla.....TrpThrSerAlaThrSerGl 2651

1602 GACTGGATTGCATACATGACAGTTCACATGGCCAGCTTTG..... 1563
      |||   |||   |||   ::::::|||   |||:::
2651 yThr...LeuGlyThrThrHisIleThrGluProSerThrGlyThrSerH 2667

1562 .....ACATGATACGTGAAGCTCCCCCCAATA 1536
                        |||      ::::::   |||   :::
2667 isThrProAlaAlaThrThrGlyThrThrThrThrSerThrProAlaLeu 2683

1535 TCGAAGAAGCACACATTACCACACCTCCTGGAAAAGATTCCATTGAAACA 1486
     |||:::           ||||||  |||::::::   :::   |||:::
2684 SerSer.....ProHis..ProSerSerArgThrThrGluSer 2695

1485 GCTCCGCATGGCATTAGCGAATGGGTGACTACTCCTGA..... 1448
        |||         |||      :::||||||||||
2696 ProPro.....SerProGlyThrThrThrProGlyHisThrThrAl 2709

1447 .....CCGAGTCGGAGAACTGTATCGA 1426
                              ||||||:::      :::|||
2709 aThrSerArgThrThrAlaThrAlaThrProSerLysThrArgThrSerT 2726

1425 AAGCTCTGGAAAGAGTTCCCTCAGCTTCTCCCTTTCCCGCGGCTTCAGAA 1376
        |||         |||:::::::::|||:::   :::::
2726 hrLeuLeuProSerGlnProThrSerAlaProIleThrThrValValThr 2742

1375 ACA..... 1373
     |||
2743 ThrGlyCysGluProGlnCysAlaTrpSerGluTrpLeuAspTyrSerTy 2759

1372 .....CATCGCACCTATGTGAGGGTT. 1352
                                    ||||||   ::::::
2759 rProMetProGlyProSerGlyGlyAspPheAspThrTyrSerAsnIleA 2776

1352 ..... 1352

2776 rgAlaAlaGlyGlyAlaValCysGluGlnProLeuGlyLeuGluCysArg 2792

1352 ..... 1352

2793 AlaThrAlaGlnProGlyValProLeuGlyGluLeuGlyGlnValValGl 2809

1351 .....TTTCGCGAATCATTGCGTC 1333
                                    ||||||::::::
2809 uCysSerLeuAspPheGlyLeuValCysArgAsnArgGluGlnValGlyL 2826

1332 CCTTTGAGAAAACATAT.....CGCACAATTTCGTGTCCAAAT 1295
        |||:::      :::           |||::::::   |||
2826 ysPheLysMetCysPheAsnTyrGluIleArgValPheCysCysAsnTyr 2842

1294 GGG.....AAAGCAATGCCGAT... 1278
     |||                    |||||||||
2843 GlyHisCysProSerThrProAlaThrSerSerThrAlaMetProSerSe 2859

1277 .....TGGTCCCTTAATACCAGTCTATTGGTGATACTCG 1244
                     |||   |||:::          |||:::
2859 rThrProGlyThrThrTrpIleLeuThrGlu......LeuThrThrThrA 2874

1243 AAACGACCGATGAATAGACCTCCTCTTGATTAAGGGCATTGGCGACGGAG 1194
        ||||||         |||:::|||              :::|||
2874 laThrThrThrAlaSerThrGlySerThrAlaThrProSerSerThrPro 2890

1193 GCCAAGATCTCCCCTCCCAACCGTGTCCTATCGCCAGCGACGATAGCGGT 1144
     :::        ||||||:::   :::   ||||||||||||      ::
2891 GlyThrAlaProProProLysValLeuThrSerProAlaThrThrProTh 2907
```

```
1143 TTTTGGGACATTGGGGAGATAGACGTCTCCGCCCTTCCACCCAAAGGTAC 1094
     :   :::|||   ::::::   ||||||            |||:::::::
2907 r...AlaThrSerSerLysAlaThrSerSerSer...SerProArgThrA 2922

1093 GCTCTTCATTG.....AGTGATGGCATG 1071
       :::::::|||                |||   :::
2922 laThrThrLeuProValLeuThrSerThrAlaThrLysSerThrAlaThr 2938

1070 TGTAGTATGTTCGTGCCCGCAAACCTAATGACCTTTAACGATCTAAAGCC 1021
                :::|||::::::  :::      :::          ||
2939 SerValThrProIleProSerSerThrLeuGlyThrThrGlyThrLeuPr 2955

1020 AAGGTGCATCAAGAAACCTATCGAAAAGGAAGATGAGAAGCAACGACGTC 971
     |:::           |||:::         :::
2955 oGluGlnThrThrThrProValAlaThrMetSerThr..... 2967

 970 TGTAAATTTCTCCCCTCAGCCTCTCATTGTTAAACCAACACCACCCTTTG 921
          |||   |||   |||   :::      :::   |||
2968 .....IleHisPro...SerSerThrProGluThrThrHisThrSerThr 2981

 920 GTGGTGCATACAAACGCCCTTTCGCGCATACTCCACAACTCAATGCGCGG 871
     |||:::   |||:::|||   :::|||           :::|||      ::
2982 ValLeuThrThrLysAlaThrThrArgAla......ThrSerSerThrSe 2996

 870 CAAGGCAACATCACACCCTGGGGCTGTATTCACGATGAAGTCCTCGATCG 821
     :      :::|||   ||||||:::::::::   :::   :::      :
2996 rThrProSerSerThrProGlyThrThrTrpIleLeuThrGluLeuThrT 3013

 820 TAGCCCTACCCCGCACAATAGGGAAGAATTTAACAGCTCTATCTTTGACG 771
     ::|||         |||   |||   :::   |||||
3013 hrAlaAlaThrThrThrAlaGlyThrGlyProThrAlaThr..... 3026

 770 TAAAAATACCCTCTCCTACGAGAGGCTTCATCCCAGATTTTAGCCCAAAC 721
           |||            :::::::::   ||||||::::::
3027 .....ProSerSerThrProGlyThrThrTrpIleLeuThrGluLe 3040

 720 GAACGTGTTCTTTACATCAATAGGAACATTA......AGGGCCCTGTTAA 677
      ::::::      |||:::   ::::::            |||   |||:
3040 uThrThrThrAlaThrThrThrAlaSerThrGlySerThrAlaThrLeuS 3057

 676 ACAATTTCCCTTTCCGC.....AAGATGACAGCG 648
     :::::   |||                  :::   ||||||
3057 erSerThrProGlyThrThrTrpIleLeuThrGluProSerThrThrAla 3073

 647 TCGCTCTCATCACCCCCACCTTTACTGAAGAAAAGGCCAAACGTCAAGGC 598
     ::::::::::::||||||                ::::::   ::
3074 ThrValThrAlaProPro.....GlySerThrAlaTh 3084

 597 GGCCTTCTCCTCCAACGACTGTTGTGGGACTGATTGCGCTGGCTCCAGTG 548
     :|||   |||::::::      :::||||||      ::::::::::::|
3084 rAlaSerSerThrGlnAlaThrAlaGlyThrProHisValSerThrThrA 3101

 547 CTTCCTGT..... 540
     ||:::
3101 laThrThrProThrValThrSerSerLysAlaThrProSerSerSerPro 3117

 539 .....TTTTTAGGAAGTGGTACGTGGACGCCAAC 511
                              |||   |||   :::   ||||||||
3118 GlyThrAlaThrAlaLeuProAlaLeuArgSerThrAlaThrThrProTh 3134

 510 AGCTGTTGGCTTAACAGCCCAATGTGACGCTTGCCGTTTGGGC..... 468
     ||||:::::::::||||||         :::      ||||||
3134 rAlaThrSerPheThrAla...IleProSerSerSerLeuGlyThrThrT 3150

 468 ..... 468

3150 rpThrArgLeuSerGlnThrThrThrProThrAlaThrMetSerThrAla 3166

 467 .....TTCTGCTCATCTTGCTTAACAGC 445
                                 |||:::   ||||||::
3167 ThrProSerSerThrProGluThrValHisThrSerThrValLeuThrTh 3183

 444 CGGTTTAGGCTGCTCGCGCTTGACGGGGGCATACGGCACTACCCGGTGGT 395
```

```
     :           :::      |||||||||   |||
  3183 r.........ThrAlaThrThrThrGlyAlaThrGly..... 3192

   394 AGTGTACGACAAGTAATGCATTAGAAACAGCTTCGAATACGACGTCGAAA 345
                       ::::::   |||   |||:::|||   :::
  3193 .................SerValAlaThrProSerSerThrProGlyThr 3203

   344 TCAGACACGTCGGAAGTTGCGGAAGTTGAGGTCCTTGCCTCCATCAAGAC 295
       :::   |||::::::|||      :::   :::   :::         ||
  3204 AlaHisThrThrLysValProThrThrThrThrThrGlyPheThrAlaTh 3220

   294 CTTCTTACCATCCCCTCTCAGCGCGCTGACTGAGCCGCTG..... 255
       |         ||||||   :::|||||||||   |||:::
  3220 rProSerSerSerProGlyThrAlaLeuThrProProValTrpIleSerT 3237

   254 ..ACAACAGCGTACTGCCGTGCAAACCCTTTAACTGGTGGGGGGACA 210
         ||||||:::         ::::::|||   |||:::|||:::|||
  3237 hrThrThrThrProThrThrThrThrProThrThrSerGlySerThr 3252

seq_name: pir1:WMTM8T

seq_documentation_block:
180K protein - tomato mosaic virus (strain L)
N;Contains: 130K protein
C;Species: tomato mosaic virus
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 30-Sep-1993
C;Accession: A04195
R;Ohno, T.; Aoyagi, M.; Yamanashi, Y.; Saito, H.; Ikawa, S.; Meshi, T.; Okada, Y.
J. Biochem. 96, 1915-1923, 1984
A;Title: Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and comp
A;Reference number: A91984; MUID:85157522
A;Accession: A04195
A;Molecule type: genomic RNA
A;Residues: 1-1615 <OHN>
A;Note: readthrough of the terminator UAG between codons for Gln-1116 and Gln-1117 occur
C;Superfamily: cucumber mosaic virus RNA 1 protein


alignment_scores:
            Quality:  309.00              Length:   1879
              Ratio:  0.445                Gaps:     64
 Percent Similarity:  36.935   Percent Identity: 16.764

alignment_block:
 US-09-301-906-3 x WMTM8T   ..

 Align seg 1/1  to: WMTM8T  from: 1  to: 1615

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||      :::   |||:::::::::::::::::::   ::
       2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::   :::|||   ::::::   |||                |||||
      18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::   |||   |||:::   ||||||            :::::|||
      35 laVal...AspGluPheAsnAlaArgAsp...........ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::   :::   :::      ::::::   ::::::
      47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
         :::   :::||||||:::::::|||   |||   ::::::   ::::::
      63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||   :::|||:::::::::||||||      ::::::   ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553

      :::         :::        :::|||||||||||||:::|||::
   97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

 1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
      :   |||:::      |||::::::      |||   |||   |||   |
  113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

 1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.....ATC 1641
      ||||||||:::|||:::   |||:::   ::::::
  130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

 1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCC..... 1680
      |||:::|||||           |||::::::::::|||
  147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

 1681 .....AGTGACCTTCTAACTGAAGCGGCTTCAA 1708
                         :::::::::   :::|||::::::
  163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

 1709 AGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC...GATTCTAGA 1755
        ::::::           |||:::||||||:::|||      :::|||
  180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196

 1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
      :::   |||      :::   ::::::|||||||||      ::::::
  197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

 1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
       :::   |||:::   :::|||::::::   |||         |||:::
  213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

 1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
        |||      :::||||||      ::::::::::|||   |||:::|||
  230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

 1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
      :::::::::      :::|||:::||||||   |||:::::::::   ::
  247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

 1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
      ::::   :::   :::   |||:::||||||:::||||||:::      :
  263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280

 2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
      ::          :::|||      :::|||:::   :::|||:::
  280 yrValCysLysThrTyrPheProAlaSerAsnArgGluValTyrMetLys 296

 2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
      |||:::
  297 GluPhe.....                                      298

 2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

  298 ..................................................  298

 2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

  298 ..................................................  298

 2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                      :::|||   |||:::   ::
  299 ..........................LeuValThrArgValAsnTh 305

 2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
      ::::        ::::::   |||   ||||||
  305 rTrpPheCysLysPheSerArgIleAspThrPhe..............  316

 2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                                  :::
  317 ....................................LeuLeuTyr 319

 2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
      :::   |||:::|||:::   ||||||
```

```
 320 LysGlyValAlaHisLysGlyValAsp..................... 328

2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446

 328 ..................................................... 328

2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
      |||   :::|||   ||||||   |||   :::
 329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis...... 341

2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
           :::   :::|||:::::::   |||   |||:::   :::
 342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
      :::      ::::::   |||:::      |||      ::::::   :
 355 uLeuGluAspSerSerSerValAsnTyrTrpProLysMetArgAspM 372

2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
     :::::   :::|||
 372 etValIleValPro..................................... 376

2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

 377 ..................................................Le 377

2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
      ||||||||||||   ||||||                |||::::::|||
 377 uPheAspIleSerLeuGluThr............SerLysArgThrArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
       :::|||:::          :::   :::|||   :::      |||:::
 390 ysGluValLeu.........ValSerLysAspPheValTyrThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   :::::::|||   ::::::|||   |||      |||   :::::
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu......ThrTyrSerAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::        |||:::|||
 418 nValLeuSerPheValGluSerIleArgSerArgVal.............. 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
           :::::::::|||      |||             :::|||:::|||
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||:::         :::   |||
 446 SerLeuLeuGlnSerLeuSerMetThr...................... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
      ||||||
 455 .PhePhe.............................................. 456

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
                                        :::::::::   :::
 457 ..................................LeuHisThrLysLeu 461

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
     ::::::|||:::   :::   :::   :::||||||   :::   |||
 462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477

3144 GAAA............GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
      |||            :::   |||:::|||:::||||||
 477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu.......... 490

3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231

 490 ..................................................... 490

3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281

 490 ..................................................... 490
```

```
3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331

 490 ..................................................... 490

3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381

 490 ..................................................... 490

3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
                                 :::|||::::::|||            ::
 491 .....................AlaPheGlyAsnAlaPheProSerIleLy 500

3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
     :      :::   :::|||:::
 500 sGluArgLeuIleAsnArgLys............................ 507

3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
                                  ::::::   |||      :::|||
 508 ........................LeuIleLysIleThrGluAsnAla 515

3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
     |||
 516 LeuGlu.................................................. 517

3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631

 518 ..................................................I 518

3632 GCGCTGTTCCCGAG......GTTGTTTGGCGAGGGAAGAGTGTCAGCTCA 3675
          ||||||:::      |||:::::::::   :::   ||||||:::
 518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534

3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
           :::   :::   :::
 535 TyrLysMetSerValAspMet............................... 541

3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
                                     |||   ::::::   |||:
 542 ..............................ProValLeuAspIleArgL 548

3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
     ::      :::||||||   :::   |||
 548 ysLysMetGluGluThrGluGluMetTyr...................... 557

3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875

 557 ..................................................... 557

3876 CGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGC 3925
                 :::::::::|||:::   :::|||:::   |||
 558 .............AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570

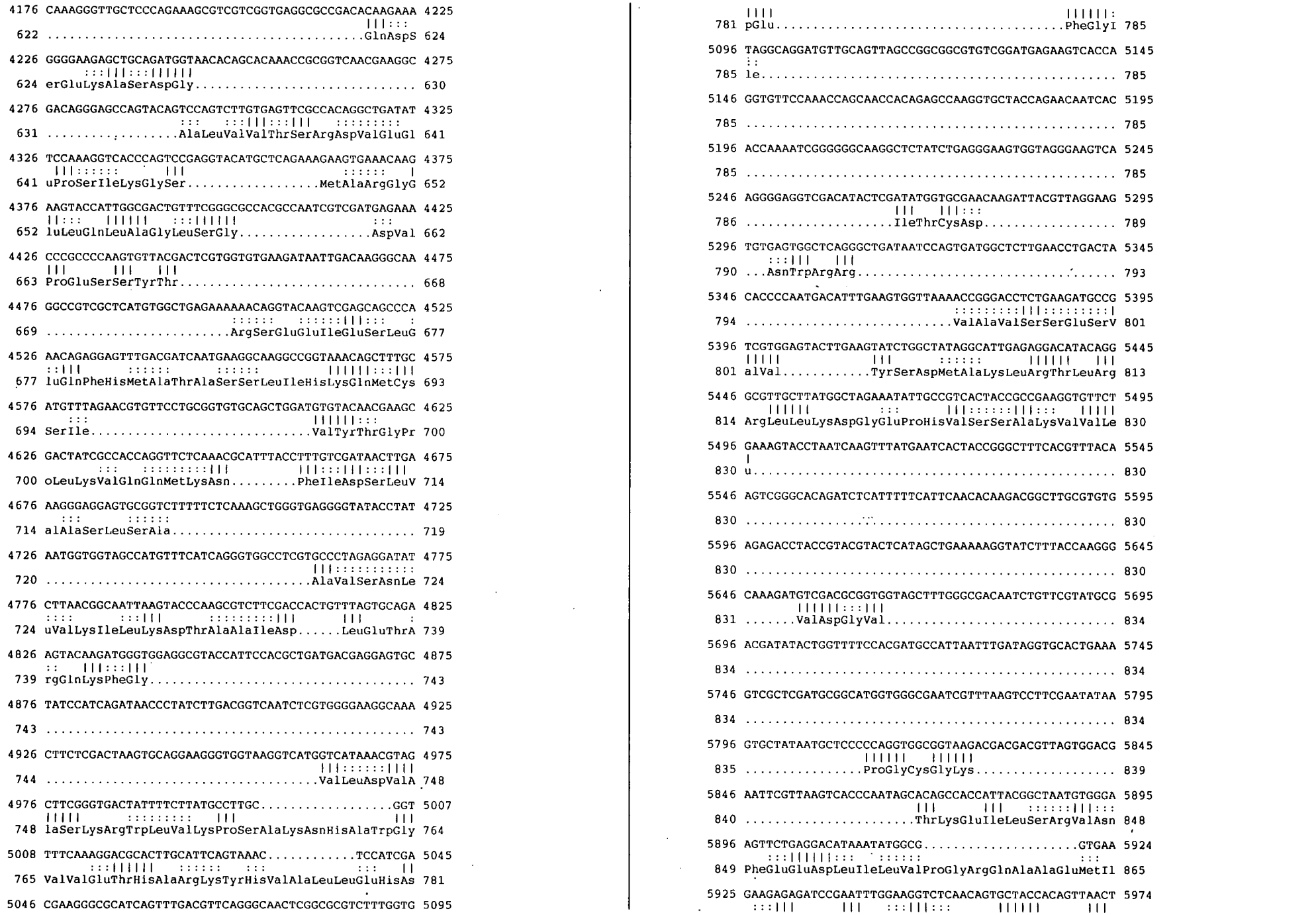
3926 GCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACC 3975
               :::::::::   :::::::::
 570 spLysPheAspValAspValPheSerGlnMetCys.............. 581

3976 CATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCC 4025
           :::|||:::                        ||||||   ||
 582 .........GlnSerLeu.....................GluValAspPr 588

4026 GTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG 4075
     |   :::|||                                      |
 588 oMetThrAla.....................................A 592

4076 CTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTC 4125
     ||   :::::::::   |||   |||   ::::::|||:::      :::
 592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuThr 608

4126 TTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGG 4175
     |||:::|||   ::::::::::::::|||      :::
 609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu.......... 621
```



```
  865 eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882

 5975 CCAGGGTGGTA...AACTTTATCGTC.................AGGGGA 6003
      |||:::|||   :::|||::::::                  |||
  882 alArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898

 6004 ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTT 6053
      ::::::||||||:::::::::|||||      |||:::|||   |||
  899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyCy 915

 6054 ACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTG 6103
      :::::::::   |||      :::   ::::::   :::      :::|
  915 sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932

 6104 GAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATG 6153
      |||||   :::|||||||||:::||||||||          |||
  932 lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948

 6154 GATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAA 6203
      :::          :::   :::::|||
  949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965

 6204 ATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA..... 6246
      :::   |||||||||   ||||||   :::::|||:::::
  965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982

 6247 ...........ATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAG 6285
              |||   :::   :::   :::|||          :::::
  982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998

 6286 GTCGTTAGCGGT......AAGGACAAACCAGTAGTAAGATCGCTGTCCAA 6329
      :::|||||||||      :::   :::||||||   :::   |||:::
  999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl 1015

 6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT 6379
       :::
 1015 yLys.........................................I 1017

 6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
      |||   :::||||||         ||||||   |||::::::|||
 1017 leLeuThrPheThrGln.........SerAsp...LysGluAlaLeuLeu 1029

 6430 GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA 6479
      ::::::|||      :::   |||   ||||||||||||:::|||||::
 1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyGl 1045

 6480 AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC 6529
      :|||::::::||||||   |||   |||          :::      |||:
 1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062

 6530 TATTCACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
      ::::::::::   :::||||||:::||||||:::|||||||||||||||
 1062 leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078

 6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
      :::|||||   |||   :::              |||   |||   ::
 1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleIl 1095

 6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
      :   |||      :::|||::::::|||:::   |||
 1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq_name: pir2:S03701

seq_documentation_block:
141K protein - pea early browning virus
C;Species: pea early browning virus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C;Accession: S03701
R;MacFarlane, S.A.; Taylor, S.C.; King, D.I.; Hughes, G.; Davies, J.W.
Nucleic Acids Res. 17, 2245-2260, 1989
A;Title: Pea early browning virus RNA1 encodes four polypeptides including a putative zi
A;Reference number: S03701; MUID:89202038
A;Accession: S03701
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1766 <MAC>
A;Cross-references: EMBL:X14006; NID:g61019; PIDN:CAB37343.1; PID:g4456716
C;Superfamily: cucumber mosaic virus RNA 1 protein


alignment_scores:
            Quality:  306.00              Length:   1922
              Ratio:  0.406                Gaps:     76
 Percent Similarity:  39.230   Percent Identity: 17.326

alignment_block:
 US-09-301-906-3 x S03701   ..

 Align seg 1/1  to: S03701  from: 1  to: 1766

 1198 GTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAG 1247
      :::::::::   ||||||   :::|||      ::::::      :::::
    7 IleSerGlnLeuLeuAsnIleAspGluPheGlyAlaGluGlnArgGlyGl 23

 1248 TATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATT 1297
      ::::   :::   ::::::   :::   :::|||   |||   :::   :
   23 nPheLeuAspLeuMetIleThrLysProAspSerGlnLeuGlyAlaMetM 40

 1298 TGGACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGC 1347
      :::::      |||   |||   :::         |||      :::|||
   40 etGlnArgIleLeuThrAspLysIle.........AspAspSerLeuArg 53

 1348 GAAAAACCCTCACATAGGTGCGATGTG.........TTTCTGAAGCCGCG 1388
      |||:::   :::::::::      |||         :::|||:::   ::
   54 GluArgLysThrArgGluThrValValIleHisGluLeuLeuSerGlnLy 70

 1389 GGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTCT 1438
      ::::::::::||||||   |||:::::::||||||:::::|||   |||:
   70 sAspGlnAsnLysLeuMetGluIleTyrProGluPheAsnIleValPheL 87

 1439 CCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTGT 1488
      ::|||:::   :::      |||   ||||||   |||   |||:::
   87 ysAspAspLysAsnMetValHisGlyPheAlaAlaAlaGluArgLysLeu 103

 1489 TTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTG...TGCTTCTTCGA 1535
              ::::::      |||      :::::            ||
  104 GlnAlaLeuLeuLeuMetAlaArgValProLysLeuGluProValAspAs 120

 1536 TATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTC 1585
      ||||||||||:::::::::::::   ::::::   |||      :::   |
  120 pIleGlyGlyGlnTrpSerPheTrpLeuSerArgGlyAspLysArgValH 137

 1586 ATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATCAAT 1635
      ||        |||::::::|||:::::::|||   :::|||:::::::::
  137 isSerSerCysProLeuIleAspMetArgAspLysGlnArgGluLeuGln 153

 1636 GAGATCCTCTTTCTT..............TCCACAGCTGGGGGAGATTC 1670
      :::      ||||||               :::::::::   ::::::||
  154 ArgGlnAsnPheLeuArgValPheArgAspAsnAlaThrThrSerGluSe 170

 1671 GTACGTGTCCAGTGACCTT......CTAACTGAAGCGGCTTCAAAGTCTG 1714
      |   :::|||:::|||         :::   :::|||   :::   ::::
  170 rArgIleSerAspAspGlnPheAspMetTyrAsnAlaPheLysGlyAspI 187

 1715 TGTCTTAC..............TGTAGTCGAGAATCGCAGAACTGCGAT 1749
      :::::                |||::::::      |||:::|||:::
  187 leAspValAlaAsnPheValArgCysAsnAsnThrPheGlnAspCysAsn 203

 1750 TCTAGAGCCGATGCGGGTTTTATGGTGGAT..................... 1779
         |||         |||:::   :::|||
  204 CysArg.........GlyTyrArgAlaAspGlyThrArgIleGlyAlaTh 217

 1780 .....................GTGTACGATATATCCCCGCAGCAGGTAG 1807
                          :::||||||:::::::   ::::::||||
  217 rAsnAlaIleAlaLeuHisSerLeuTyrAspPheLysLeuAspAspValA 234
```

```
1808 CAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATG 1857
     ||:::||||||   :::||||||:::      :::      |||:::::: 
 234 laAspAlaMetIleGluLysGlyThrLysPheLeuHisAlaAlaMetLeu 250

1858 TTCCCCGTGGAGTTG.....................TTGTACGGTAACGG 1886
     |||      |||                        |||   :::      
 251 PheAlaProGluAlaGluIleGluLysGluGlyProLeuProSerValAs 267

1887 TGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAGGTGATT 1936
            |||   ::::::      :::|||::::::   |||      :
 267 pGlyTyrTyrGluArgLysGluGlySerLeuIleSerSerGluLysIleP 284

1937 ACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTC 1986
     ::::::::::|||   ::::::   :::      |||   |||:::::: 
 284 hePheGlyPheAsn...AsnAspProSerTyrAlaTyrIleHisAspTrp 299

1987 TCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGC........... 2025
     |||:::   :::               :::|||:::||| 
 300 SerGluTyrLys.............LysTyrLeuArgGlyGluProPh 311

2026 .ACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTG 2074
      :::      |||:::||||||   :::|||             |||  
 311 eSerArgArgGlyHisValPheMetPheGluProTrpGlnAlaArgGlyA 328

2075 GTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACTGAG 2124
                :::|||:::   |||             |||   |||::: 
 328 spThrMetPhePheThrLeuTyrArgMetThrGlyValProMetThrAsn 344

2125 GTTACG............TATAGGTCGTTG...GTCCCGTCGTTCGTGGG 2159
     :::               ||||||   |||   :::      :::   || 
 345 LeuLeuGlyAsnGluTyrTyrArgArgLeuHisIleSerArgTrpGluGl 361

2160 CAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTA 2209
     |      :::|||   :::||||||         ::::::::::::   | 
 361 y......MetValValValProValPheGluIleAspGluIleThr...L 375

2210 AGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCG 2259
     ||   ::::::   :::|||   :::|||::::::   ||| 
 375 ysLysLeuThrLysSerSerMetTyrValGluLysAlaTyrMet...... 389

2260 CTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGG 2309
                           ::::::   ::::::|||::: 
 390 .......................AspLysCysLeuAspTyrVal..... 396

2310 GGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCC 2359
                                              |||||||::: 
 397 .....................................SerArgLeuSerA 401

2360 ACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCT 2409
       :::::::::   ||| 
 401 spGlnGlnLeuThrIle................................ 406

2410 GTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGT...ATTCGCTCCTA 2456
                           ::::::::::||||||   :::   ||| 
 407 ......................AsnAsnValLysSerPheMetSerSerAs 416

2457 TAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCT 2506
      |||::::::
 416 nAsnTrpVal..................................... 419

2507 TCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCT 2556
                             |||   |||:::::::::|||
 420 ....................LeuPheIleAsnGlyAlaAlaValLysAsn 429

2557 AAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGCTTATGTCTAT 2606
     |||      |||
 430 LysGlnSerValAsp................................. 434

2607 GCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTT 2656
      ||||||
 435 .ProArg..................................... 436
```

```
2657 GCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATT 2706
                              ||||||:::|||   :::          ::: 
 437 ....................ArgLeuGlnLeuLeuAlaGlnThrLeuLeu 446

2707 TCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTT 2756
       ||||||
 447 ValLysGlu..................................... 449

2757 CTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGA 2806
                                     |||:::|||:::   :::   | 
 450 ............................LysLeuMetArgProLeuMetT 457

2807 CAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTC 2856
     |||||      :::   :::   :::
 457 hrGluMetArgGluLysMetValLeu..................... 465

2857 ATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTGG 2906

 465 ............................................... 465

2907 CGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTG...GTAG 2953
            :::||||||:::|||:::||||||:::      |||   |||| 
 466 .......ArgAlaThrAlaValAsnSerValThrGlyThrLeuLysValA 480

2954 CTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCC 3003
     ||   |||                                 ::::::|||::: 
 480 laTrpGly............................AsnTyrPheAsp 486

3004 ACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGA 3053
        |||||||||                                 |||
 487 ...GlySerLeu............................ArgArgLy 492

3054 TGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGG 3103
      :::|||::::::      :::::::::::::   :::|||::: 
 492 sCysLeuLeuLysLeuAlaGlnValLeuGlyThrGluThrGlyLeu.... 507

3104 AGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTT 3153
                                        |||      :::|||::: 
 508 ........................GluValLeuGluValLysAspAla 515

3154 TGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAA 3203
        :::   :::::::::||||||:::   |||·  ::: 
 516 ProLysTyrIleGluIleAsnAspTyrLeuThrAlaIle........... 528

3204 GGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCAT 3253

 528 ............................................... 528

3254 GGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGT 3303

 528 ............................................... 528

3304 AGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACG 3353
                  :::::::::   ::::::::::::   |||:::||| 
 529 ............PheAsnGluThrThrGluLysThrAspGlySerLeuPr 541

3354 CGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAG 3403
        |||   ::::::                          ::: 
 541 oAspLeuAspGluAla............................LysA 548

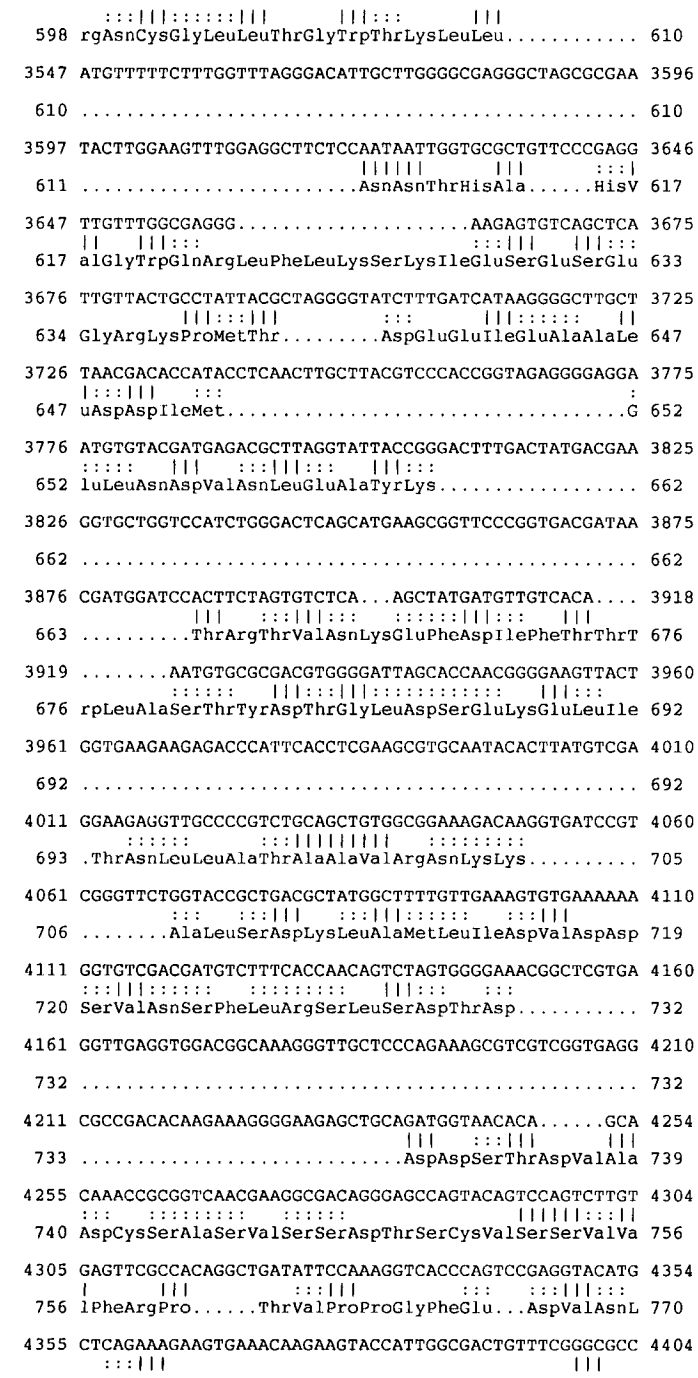
3404 GGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAA 3453
        |||:::   ::::::          |||   :::|||   ::: 
 548 rgAsnSerAspLysIleSerLysGluAlaAlaGluAlaAlaValGlnCys 564

3454 GTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCC..... 3498
     |||   :::   :::   :::|||::::::|||   :::::::::
 565 ValLysMetGlnPheProLysPheGluSerSerAspSerLeuLysGluPr 581

3498 ............................................... 3498

 581 oLeuIleArgLysIleAspGlyAspValLysLysLysGlyLysGlnArgA 598

3499 ..GGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACTTA 3546
```


598 rgAsnCysGlyLeuLeuThrGlyTrpThrLysLeuLeu 610
3547 ATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAA 3596
610 610
3597 TACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAGG 3646
611 AsnAsnThrHisAla HisV 617
3647 TTGTTTGGCGAGGG AAGAGTGTCAGCTCA 3675
617 alGlyTrpGlnArgLeuPheLeuLysSerLysIleGluSerGluSerGlu 633
3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
634 GlyArgLysProMetThr AspGluGluIleGluAlaAlaLe 647
3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
647 uAspAspIleMet G 652
3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
652 luLeuAsnAspValAsnLeuGluAlaTyrLys 662
3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875
662 662
3876 CGATGGATCCACTTCTAGTGTCTCA AGCTATGATGTTGTCACA 3918
663 ThrArgThrValAsnLysGluPheAspIlePheThrThrT 676
3919 AATGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACT 3960
676 rpLeuAlaSerThrTyrAspThrGlyLeuAspSerGluLysGluLeuIle 692
3961 GGTGAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGA 4010
692 692
4011 GGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGT 4060
693 ThrAsnLeuLeuAlaThrAlaAlaValArgAsnLysLys 705
4061 CGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAA 4110
706 AlaLeuSerAspLysLeuAlaMetLeuIleAspValAspAsp 719
4111 GGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGA 4160
720 SerValAsnSerPheLeuArgSerLeuSerAspThrAsp 732
4161 GGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGG 4210
732 732
4211 CGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACA GCA 4254
733 AspAspSerThrAspValAla 739
4255 CAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGT 4304
740 AspCysSerAlaSerValSerSerAspThrSerCysValSerSerValVa 756
4305 GAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATG 4354
756 lPheArgPro ThrValProProGlyPheGlu AspValAsnL 770
4355 CTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCC 4404


770 euArgLys GlyLys 774
4405 ACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGT 4454
775 ThrValIleValAspAsnAspAlaAlaGluSerSerSerSer 788
4455 GAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAAC 4504
789 SerGluArgAsnA 793
4505 AGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGC 4554
793 rg 793
4555 AAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCA 4604
793 793
4605 GCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCAT 4654
794 AsnHisPheAlaAsn P 799
4655 TTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAG 4704
799 heGluValIleGluAsn CysArg 806
4705 CTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTG 4754
807 PheGlyAspAlaProLysGluThrGlyAspPheSerValAspSerArg 822
4755 GCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCT 4804
823 LeuGluPheIleHisTyrLeu 829
4805 TCGACCACTGTTTAGTG 4821
830 ArgCysLeuIleCysAlaGlnAsnAsnGluLeuLeuGlyLysTyr 844
4822 CAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGA 4871
845 ArgAspTyrGluMetGly 850
4872 GTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGG 4921
850 850
4922 CAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAG 4956
851 ValValArgProGlyGlyLysGlyTyrProAspGlu 862
4957 GTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCT 5001
863 LeuGlyValPheAspLeuAlaLeuLysLysTrpIleIleLysProProSe 879
5002 TGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAG 5050
879 rCysSerTyrAsnLysAlaPheValProAspValSerAlaLysGluGlnG 896
5051 GGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGC 5100
896 lyLys TrpIleGly 900
5101 AGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGT 5150
901 TyrLeuValAspAlaSerTrpGly 908
5151 TCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAA 5200
909 L 909
5201 AATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGG 5250
909 ysGlnGlyIleAspAlaPheSer 916

```
5251 AGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGA 5300
                                                    |||
 917 ..............................................Cys.. 917

5301 GTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCC 5350
                                                  ||||||
 918 ..........................................TyrThrA 920

5351 CAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTG 5400
       ::::::::::::::
 920 snValAlaTrpLysAla.............................. 925

5401 GAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTT 5450

 925 ..................................................... 925

5451 GCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAG 5500
                  :::||||||::::::   ::::::
 926 .............AspIleAlaIleValCysSerGln..... 933

5501 TACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCG 5550

 933 ..................................................... 933

5551 GGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGA 5600

 933 ..................................................... 933

5601 CCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAG 5650

 933 ..................................................... 933

5651 ATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGAT 5700
                                          |||:::|||::::::
 934 ..............................ThrPheLeuCysAsnGlu 939

5701 ATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGC 5750
       :::::::::          ||||||   |||   |||:::|||::
 940 ArgIleIleLeuLys.........AsnLeuAlaGly...LeuGluValVa 952

5751 T......CGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATA 5794
     :      ||||||                                ::::::|
 952 lProLeuArgCys..............................LysPheL 959

5795 AGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGAC 5844
     ||      ::::::   ||||||   |||||||:::|||   :::|||:::
 959 ysLeuValAspGlyValProGlyCysGlyLysSerThrMetIleValAsn 975

5845 GAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGG 5894
         :::   :::|||       ::::::::::::::::
 976 ......ThrAlaAsnProValPheAspValValLeuSerThrCysLysGl 990

5895 AAGTTCTGAGGACATA......AATATGGCGGTGAAGAAGAGAGATCCGA 5938
      ::::::||||||:::      :::::::|||:::||||||         |
 990 uAlaThrGluAspLeuLeuGluLysPheAlaAlaLysLysPheGlyIleA 1007

5939 ATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAAC 5988
     ||||||:::      :::   :::   ||||||:::|||   :::::::::
1007 snLeuLys......LysArgValLysThrValAspSerPheLeuMetHis 1021

5989 TTTATCGTCAGGGGAATGTATAAAAGGGTTTTG....GTGGATGAGGTGC 6034
       |||:::|||:::|||:::      :::         ||||||:::
1022 .CysSerAspGlyAsnCysValGlyAspLeuLeuHisPheAspGluAlaL 1038

6035 ACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCG 6084
       |||   |||  |||::::::   :::   :::      :::   |||
1038 euMetAlaHisAlaGlyMetValPhePheCysAlaGlnIleAlaLysAla 1054

6085 TCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAG 6134
     ::::::   :::      ||||||   :::||||||   :::       ||
1055 LysLysValIleCysGlnGlyAspGlnLysGlnIleAlaTyrLysProAr 1071

6135 GGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAA 6184
     |   :::   :::   :::      :::   ::::::   :::   ::::
1071 gValSerGlnLeuThrLeuArgPheThrSerLeuIleGlyArgPheAspG 1088

6185 GCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTAC 6234
     ::|||              ||||||||||||||:::|||||||:::
1088 luValGluGluLysArgMetSerTyrArgCysProValAspValAlaLeu 1104

6235 TTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAA 6284
       |||                          ::::::   |||      ||
1105 ThrLeu.....................AspArgPheTyrThrGlyLy 1113

6285 GGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGC 6334
     |||||||:::         :::   |||:::||||||::::::   :::
1113 sValValThr......LysAsnSerValLeuArgSerMetAspValLysA 1128

6335 CAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTG 6384
       |||||||:::   ::::::|||         ::::::      ||||||
1128 rgIleGlySerLysGluGlnValGluMetGluHisGlyIleGlnTyrLeu 1144

6385 TGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAA 6434
       :::   |||   ||||||:::|||:::   :::   |||      ::
1145 ThrPheLeuGlnSerGluLysLysAspIleAlaAsnLeuLeuCysGlnAr 1161

6435 AGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACAT 6484
     :      ::::::   |||   ||||||||||||||||||||||||||||
1161 gLysValLysSerPheValAsnThrValHisGluAlaGlnGlyLysThrP 1178

6485 TCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTC 6534
     ||:::   |||   |||   |||   |||   :::||||||   |||
1178 heLysLysValArgLeuValArgLeuLysProThrAspAspValLeuAla 1194

6535 ACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACT 6584
           |||   :::   :::|||:::||||||||||||||:::|||||
1195 ArgGlyGlnGluTyrGlnIleValAlaLeuSerArgHisThrGlnSerLe 1211

6585 GGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATA 6634
     |||||||   :::::::::          ||||||:::|||::::::   |
1211 uValTyrGluThrValLys.........AspAspGluValSerAlaLeuI 1225

6635 TTAGCGACGCGTCA 6648
     ||   |||::::::
1225 leArgAspSerAla 1229
```

seq_name: pir1:WMTMPV

seq_documentation_block:
183K protein - pepper mild mottle virus (strain Spain)
N;Contains: 126K protein
C;Species: pepper mild mottle virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994
C;Accession: JQ1312
R;Alonso, E.; Garcia-Luque, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra,
J. Gen. Virol. 72, 2875-2884, 1991
A;Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resist
A;Reference number: JQ1312; MUID:92113528
A;Accession: JQ1312
A;Molecule type: genomic RNA
A;Residues: 1-1611 <ALO>
A;Cross-references: GB:M81413
A;Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln
C;Comment: This protein may have RNA polymerase activity.
C;Superfamily: cucumber mosaic virus RNA 1 protein
F;1-1117/Product: 126K protein #status predicted <PRO>

alignment_scores:
Quality: 297.00 Length: 854
Ratio: 0.754 Gaps: 40
Percent Similarity: 46.136 Percent Identity: 22.834

alignment_block:
US-09-301-906-3 x WMTMPV ..

```
Align seg 1/1  to: WMTMPV  from: 1  to: 1611

4522 CCCAAACAGAGGAGTTTGACGATC..........................  4545
     ||||||:::::::::::::::|||
 366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuGl  382

4546 .AATGAAGGCAAG......GCCGGTAAACAGCTTTGCATGTTTAGAACGT  4588
      ||||||||||||      |||   |||:::::   :::   :::
 382 nAsnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAspP  399

4589 GTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACC  4638
           |||       :::   :::      :::   |||
 399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu  415

4639 AGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGG.....  4683
      ::::::|||:::::::::|||||||:::::::::::::::|||
 416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValIl  432

4683 .................................................  4683

 432 eIleAsnGlyValThrAlaArgSerGluTrpAspValAspLysAlaLeuL  449

4684 ...........AGTGCGGTCTTTTTC......TCAAAGCTGGGT......  4710
             |||   :::|||||      :::|||||||:::
 449 euGlnSerLeuSerMetThrPhePheLeuGlnThrLysLeuAlaMetLeu  465

4711 ................GAGGGGTATACCTATAATGGTGGTAGCCATGT  4742
                     :::   :::      :::::   |||   ::
 466 LysAspAspLeuValValGlnLysPheGlnValHisSerLysSerLeuTh  482

4743 TTCATCAGGGTGG.............................CCTC  4759
     ::::      |||                              |||
 482 rGluTyrValTrpAspGluIleThrAlaAlaPheHisAsnCysPheProT  499

4760 GTGCCCTAGAGGATATCTTAACG.......................  4782
           |||   :::::::::
 499 hrIleLysGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys  515

4783 ......GCAATTAAGTACCCAAGCGTC.........TTCGACCACTGTTT  4817
           ||||||   |||:::::         |||         ||
 516 AlaLeuGluIleLysValProAspLeuTyrValThrPheHisAspArgLe  532

4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCA.........TTCCACG  4858
     ||||:::::||||||   :::      :::|||
 532 uValLysGluTyrLysSerServalGluMetProValLeuAspValLysL  549

4859 CTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAAT  4908
       :::   ||||||:::           |||   :::      ::::::
 549 ysSerLeuGluGluAlaGluValMetTyrAsnAlaLeuSerGluIleSer  565

4909 CTCGTGGGGAAGGCAAAC.................TTCTCGACTAAGTG  4940
     ::::::      ::::::                 ||||||      ||
 566 IleLeuLysAspSerAspLysPheAspValAspValPheSerArgMetCy  582

4941 CAGGAAGGGTGGT.....................AAGGTCATGGTCATAA  4969
     |:::      |||                     ||||||||||||
 582 sAsnThrLeuGlyValAspProLeuValAlaAlaLysValMetValAlaV  599

4970 ACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACG  5019
       |||:::
 599 alValSer.....................................  601

5020 CACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTT  5069
                              :::|||:::   ::::::|||||||
 602 ......................AsnGluSerGlyLeuThrLeuThrPh  610

5070 CAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG  5119
     |:::         ::::::   :::|||:::      ||||||
 610 eGluArgProThrGluAla...AsnValAlaLeuAlaLeuGln.......  623

5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAG  5169
                                      |||:::         :::

 624 ...............................ProThrIleThrSerLys  629

5170 AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCT  5219
     ::::::|||:::   :::   ::::::         |||
 630 GluGluGlySerLeuLysIleValSerSerAspValGly...........  642

5220 ATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATA.  5268
       |||:::|||   :::||||||      :::|||      |||:::
 643 ....GluSerSerIleLysGluValValArgLysSerGluIleSerMetL  658

5269 ................TGGTGCGAACAAGATTACGTTAGGAAGTGTGAG  5301
                                 ::::::::::::   |||:::   |||
 658 euGlyLeuThrGlyAsnThrValSerAspGluPheGlnArgSerThrGlu  674

5302 TGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCC  5351
                           :::   :::|||:::
 675 ....................IleGluSerLeuGln..............  679

5352 AATGACATTTGAAGTGGTTAAAACCGGGACC.........TCTGAAGATG  5392
           |||::::::|||:::|||   |||         ::::::
 680 ....GlnPheHisMetValSerThrGluThrIleIleArgLysGlnMetH  695

5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC  5442
       ::::::   |||         |||   :::         ::::::|||
 695 isAlaMetValTyrThrGlyProLeuLysValGlnGlnCysLysAsnTyr  711

5443 ...AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGG  5489
           :::|||:::   :::   :::   ||||||::::::   :::
 712 LeuAspSerLeuValAlaSerLeuSerAlaAlaValSerAsnLeuLysLy  728

5490 TGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACG  5539
      ::::::|||
 728 sIleIleLys.....................................  731

5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTG  5589
                    |||      |||      :::|||::::::   :::
 732 ...........AspThrAlaAlaIleAspLeuGluThrLysGluLysPhe  744

5590 CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATC.....  5634
       |||            |||   :::      :::|||   :::
 745 GlyVal...........TyrAspValCysLeuLysLysTrpLeuValLy  757

5635 ....TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACA  5680
       ::::::||||||           :::|||:::          |||:
 757 sProLeuSerLysGlyHisAlaTrpGlyValValMet.........AspS  771

5681 ATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTG  5730
     ::   :::   |||         |||:::   :::|||:::   |||:::
 771 erAspTyrLysCysPheValAlaLeuLeuThrTyrAspGlyGluAsnIle  787

5731 ATAGGTGCACTGAAAGTCGCTCGATGCGGC............ATGGTGGG  5768
     :::                     ||||||               |||::
 788 Val...................CysGlyGluThrTrpArgArgValAl  797

5769 CGAATCGTTTAAGTCCTTCGAATATAAGTGCTAT.............  5802
     :   |||   :::|||:::   |||:::
 797 aValSerSerGluSerLeuValTyrSerAspMetGlyLysIleArgAlaI  814

5803 ...................AATGCTCCCCCAGGTGGCGGTAAGACGACG  5832
                        :::   |||      ::::::      :::
 814 leArgSerValLeuLysAspGlyGluProHisIleSerSerAlaLysVal  830

5833 ACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTAC  5882
     ||||||||||||      |||   :::   ::::::|||      |||
 831 ThrLeuValAspGly...ValProGlyCysGlyLysThrLysGluIleLe  846

5883 GGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAG....  5928
      ::::::|||:::   :::||||||:::   ::::::      |||
 846 uSerArgValAsnPheAspGluAspLeuValLeuValProGlyLysGlnA  863

5929 ................AGAGATCCGAATTTGGAAGGTCTCAACAGTGCT  5961
                     |||      |||   :::||||||      |||
 863 laAlaGluMetIleArgArgArgAlaAsnSerSerGlyLeuIleValAla  879
```

```
 5962 ACCACAGTTAACTCCAGGGTGGTA...AACTTTATCGTC.........AG 5999
      |||      |||   |||:::|||   :::|||::::::         ||
  880 ThrLysGluAsnValArgThrValAspSerPheLeuMetAsnTyrGlyAr 896

 6000 GGGA......ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGC 6043
      ||||      :::|||||||||:::::::::||||||      |||:::|
  896 gGlyProCysGlnTyrLysArgLeuPheLeuAspGluGlyLeuMetLeuH 913

 6044 ATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGC 6093
      ||   |||   :::::::::   |||      :::   :::||||||:::
  913 isProGlyCysValAsnPheLeuValGlyMetSerLeuCysSerGluAla 929

 6094 CTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGT 6143
      :::   :::||||||   :::|||||||||:::|||||||||      ::
  930 PheValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValAlaTh 946

 6144 GTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTAT 6193
      :|||              :::         :::   :::::|||
  946 rPheProTyrProLysHisLeuSerGlnLeuGluValAspAlaValGluT 963

 6194 ACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
                :::   |||||||||   |||:::   ::::::|||:::
  963 hrArgArgThrThrLeuArgCysProAlaAspIleThrPhePheLeuAsn 979

 6244 TCA.....ATGACCGTAAGGGGAACGGAAAAGTGTTA 6275
      :::      |||   :::   ::::::   :::
  980 GlnLysTyrGluGlyGlnValMetCysThrSerSerValThrArgSerVa 996

 6276 CCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGT 6325
          ::::::|||:::::|||           |||:::::   :::|
  996 lSerHisGluValIleGlnGly......AlaAlaValMetAsnProValS 1011

 6326 CCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGAC 6375
      |||||        |||   :::
 1011 erLysProLeuLysGlyLysVal..........................  1018

 6376 GTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTT 6425
           :::   :::||||||   :::||||||                ||
 1019 ......IleThrPheThrGlnSerAspLysSer.............LeuLe 1029

 6426 GAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGG 6475
      |   ::::::|||   |||      |||   ||||||||||||:::||||
 1029 uLeuSerArgGlyTyrGlu...AspValHisThrValHisGluValGlnG 1045

 6476 GAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGAC 6525
      ||:::|||||||:::||||||   |||   |||         :::
 1045 lyGluThrPheGluAspValSerLeuValArgLeuThrProThrProVal 1061

 6526 TCCCTATTCACTAAACAA...CCGCATATACTTGTTGGTTTGTCGAGACA 6572
      ::::::::::::||||||   ||||||:::||||||:::||||||||||||
 1062 GlyIleIleSerLysGlnSerProHisLeuLeuValSerLeuSerArgHi 1078

 6573 CACACGCTCACTGGTTTAT..............GCCGCTCTGAGCTCAA 6607
      ||||||||||:::   |||                |||:::   |||
 1078 sThrArgSerIleLysTyrTyrThrValValLeuAspAlaValValSerV 1095

 6608 AGTTGGACGATAAG......GTCGGCACATATATTAGCGACGCGTCACCT 6651
        |||   |||         |||::::::|||:::   |||
 1095 alLeuArgAspLeuGluCysValSerSerTyrLeuLeuAspMetTyrLys 1111

 6652 CAATCAGTATCC 6663
        :::||||||
 1112 ValAspValSer 1115

seq_name: pir1:WMTMS2

seq_documentation_block:
186K protein - cucumber green mottle mosaic virus (strain SH)
C;Species: cucumber green mottle mosaic virus
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Apr-1994
C;Accession: JQ1157
R;Ugaki, M.; Tomiyama, M.; Kakutani, T.; Hidaka, S.; Kiguchi, T.; Nagata, R.; Sato, T.;
J. Gen. Virol. 72, 1487-1495, 1991
A;Title: The complete nucleotide sequence of cucumber green mottle mosaic virus (SH s
A;Reference number: JQ1157; MUID:91311400
A;Accession: JQ1157
A;Molecule type: genomic RNA
A;Residues: 1-1646 <UGA>
A;Cross-references: GB:D12505
C;Superfamily: cucumber mosaic virus RNA 1 protein


alignment_scores:
            Quality:  294.50              Length:   1919
              Ratio:  0.431                Gaps:     70
 Percent Similarity:  35.591   Percent Identity: 17.040

alignment_block:
 US-09-301-906-3 x WMTMS2   ..

 Align seg 1/1  to: WMTMS2  from: 1  to: 1646

 1213 AATCAAGAGGAGGTCTATTCATCG.............GTCGTTTCGAGTAT 1250
      ||||||:::::::::   :::|||             :::||||||:::::
   10 AsnGlnArgAspAlaAlaAlaSerGlyArgAsnAsnLeuValSerGlnLe 26

 1251 CACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGG 1300
      ::::::::::   |||      :::              |||:::|||   |
   26 uAlaSerLysArgValTyrAspGluAlaValArgSerLeuAspHisGlnA 43

 1301 ACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAA 1350
      ||                                              :::
   43 sp.............................................Arg 44

 1351 AAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAA 1400
      :::|||:::   :::   :::      ::::::   ::::::   :::
   45 ArgProLysMetAsnPheSerArgValValSerThrGluHisThrArgLe 61

 1401 GCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCA 1450
       :::   :::   :::||||||:::||||||:::|||:::   :::   :
   61 uValThrAspAlaTyrProGluPheSerIleSerPheThrAlaThrLysA 78

 1451 GGAGTAGTCACCCATTCGCTAATGCCATGCCGGAGCTGTTTCAATGGAATC 1500
      ::|||   |||   :::|||:::::::::|||
   78 snSerValHisSerLeuAlaGlyGlyLeuArgLeuLeuGluLeuGluTyr 94

 1501 TTTTCCAGGAGGTGT......GGTAATGTGTGCTTCTTCGATATTGGGGG 1544
      :::      :::         |||:::   |||   :::||||||||||
   95 MetMetMetGlnValProTyrGlySerProCys...TyrAspIleGlyGl 110

 1545 GAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCA 1594
      |:::::|||   |||:::      |||:::         |||   ||||
  110 yAsnTyrThrGlnHisLeuPheLysGlyArgSerTyrValHisCysCysA 127

 1595 ATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATC......AATGAG 1638
      |||||   |||||:::||||||||||   |||:::::      |||:::
  127 snProCysLeuAspLeuLysAspValAlaArgAsnValMetTyrAsnAsp 143

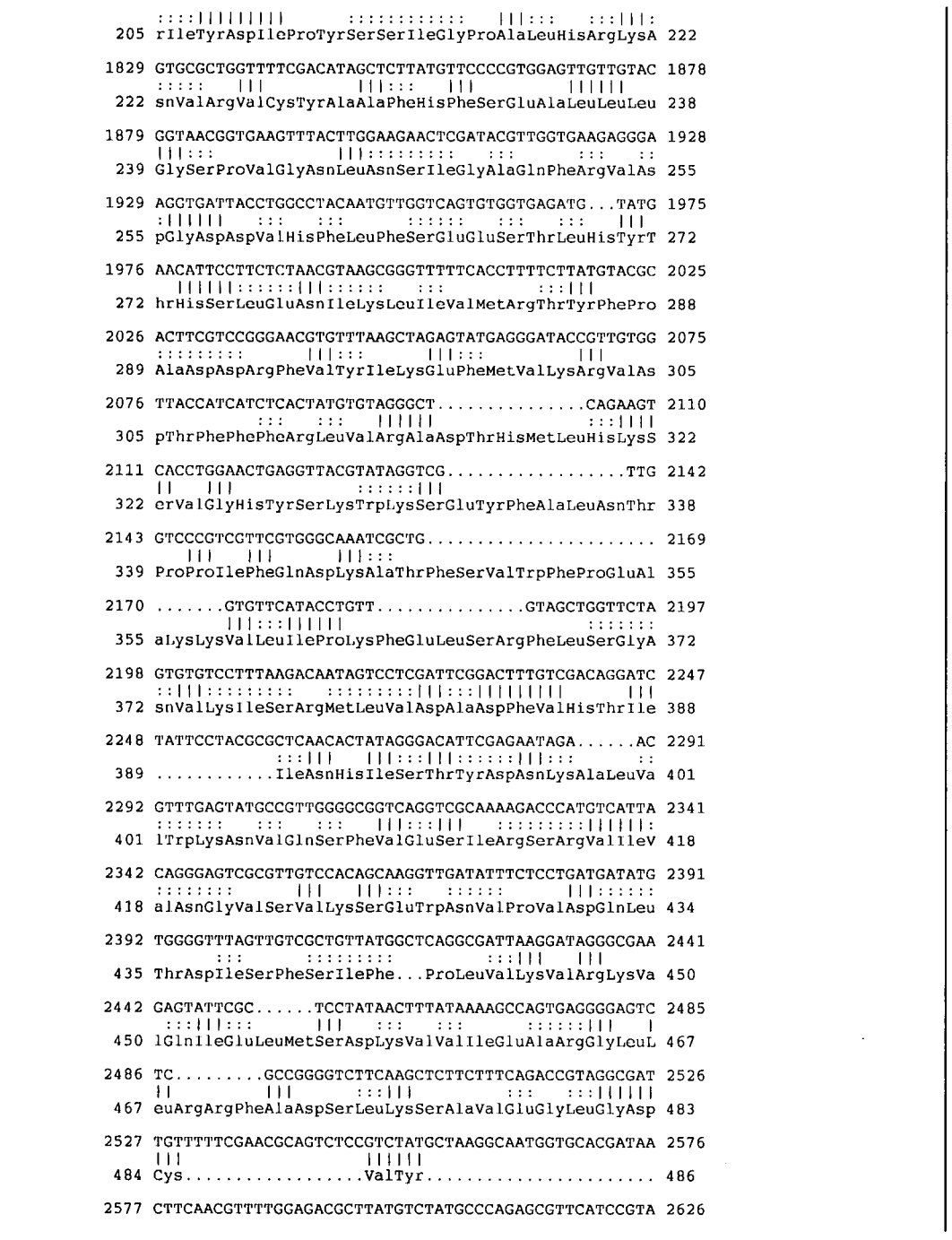
 1639 ATCCTCTTT..................................... 1647
      ::::::
  144 MetIleThrGlnHisValGlnArgHisLysGlySerCysGlyCysArgPr 160

 1648 .CTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTG 1696
       |||   |||        |||:::::      |||
  160 oLeuProThrPheGlnIleAspAlaPheArgArgTyrAsp..........  173

 1697 AAGCGGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC 1746
           :::|||      :::      ||||||         |||:::|||
  174 .....SerSerProCysAlaValThrCysSerAspValPheGlnGluCys 188

 1747 GATTCTAGAGCCGATGCGGGT.................TTTATGGTGGA 1778
      :::           :::|||                    :::   ::
  189 SerTyrAspPheGlySerGlyArgAspAsnHisAlaValSerLeuHisSe 205

 1779 TGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGG 1828
```


205 rIleTyrAspIleProTyrSerSerIleGlyProAlaLeuHisArgLysA 222
1829 GTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTAC 1878
222 snValArgValCysTyrAlaAlaPheHisPheSerGluAlaLeuLeuLeu 238
1879 GGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGA 1928
239 GlySerProValGlyAsnLeuAsnSerIleGlyAlaGlnPheArgValAs 255
1929 AGGTGATTACCTGGCCCTACAATGTTGGTCAGTGTGGTGAGATG...TATG 1975
255 pGlyAspAspValHisPheLeuPheSerGluGluSerThrLeuHisTyrT 272
1976 AACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGC 2025
272 hrHisSerLeuGluAsnIleLysLeuIleValMetArgThrTyrPhePro 288
2026 ACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGG 2075
289 AlaAspAspArgPheValTyrIleLysGluPheMetValLysArgValAs 305
2076 TTACCATCATCTCACTATGTGTAGGGCT.....CAGAAGT 2110
305 pThrPhePhePheArgLeuValArgAlaAspThrHisMetLeuHisLysS 322
2111 CACCTGGAACTGAGGTTACGTATAGGTCG.....TTG 2142
322 erValGlyHisTyrSerLysTrpLysSerGluTyrPheAlaLeuAsnThr 338
2143 GTCCCGTCGTTCGTGGGCAAATCGCTG..... 2169
339 ProProIlePheGlnAspLysAlaThrPheSerValTrpPheProGluAl 355
2170 GTGTTCATACCTGTT.....GTAGCTGGTTCTA 2197
355 aLysLysValLeuIleProLysPheGluLeuSerArgPheLeuSerGlyA 372
2198 GTGTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATC 2247
372 snValLysIleSerArgMetLeuValAspAlaAspPheValHisThrIle 388
2248 TATTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGA.....AC 2291
389 IleAsnHisIleSerThrTyrAspAsnLysAlaLeuVa 401
2292 GTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTA 2341
401 lTrpLysAsnValGlnSerPheValGluSerIleArgSerArgValIleV 418
2342 CAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATG 2391
418 alAsnGlyValSerValLysSerGluTrpAsnValProValAspGlnLeu 434
2392 TGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAA 2441
435 ThrAspIleSerPheSerIlePhe...ProLeuValLysValArgLysVa 450
2442 GAGTATTCGC.....TCCTATAACTTTATAAAAGCCAGTGAGGGGAGTC 2485
450 lGlnIleGluLeuMetSerAspLysValValIleGluAlaArgGlyLeuL 467
2486 TC.....GCCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGAT 2526
467 euArgArgPheAlaAspSerLeuLysSerAlaValGluGlyLeuGlyAsp 483
2527 TGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAA 2576
484 Cys.....ValTyr..... 486
2577 CTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTA 2626
486 486
2627 AAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGAC 2676
487 Asp 487
2677 AGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAG 2726
488 AlaLeuValGlnThrGlyTrpPheAspThrSer..... 498
2727 GAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAG 2776
499 SerA 500
2777 ATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTG 2826
500 spGluLeuLysValLeu.....LeuProGlu 508
2827 CCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAA 2876
509 ProPheMetThrPheSerAspTyrLeu..... 517
2877 CGTCCACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGG 2926
518 GluGlyMetTyrGluAlaAspAlaLysIleGluA 529
2927 TTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCC 2976
529 rgGluSerValSerGluLeuLeuAlaSerGlyAspAspLeuPheLysLys 545
2977 TTTTCTGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCG 3026
546 IleAspGluIleArgAsnAsnTyrSer..... 554
3027 CGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATG 3076
555 .GlyValGluPheAsp..... 559
3077 CTATAAACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGC 3126
559 559
3127 AAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGA 3176
560 ValGluLysPheGlnGluPheCysLysGluLeuAsnVal..... 572
3177 CTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGG 3226
573 AsnProMetLeuIleGly..... 578
3227 CGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCG 3276
578 578
3277 GCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTT 3326
579 HisValIleGluAlaIl 584
3327 AAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTAT 3376
584 ePheSerGlnLysAlaGly..... 590
3377 CCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTC 3426
591 ValThrVal 593
3427 ACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTT 3476
594 Thr..... 594
3477 TTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCAC 3526
594 594

```
3527 ATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTT 3576
                                      ||||||||||||||||
 595 ..............................GlyLeuGlyThrLeu... 599

3577 GGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAA 3626

 599 ..................................................... 599

3627 TTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCAT 3676
               ||||||:::          |||   ||||||:::
 600 ..........SerProGluMet.........GlyAlaSerValAla.... 608

3677 TGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTT 3726
                                                     |||
 609 ..................................................Leu 609

3727 AACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAA 3776
     ::::::|||
 610 SerSerThr.......................................... 612

3777 TGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAG 3826

 612 ..................................................... 612

3827 GTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAAC 3876

 612 ..................................................... 612

3877 GATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCG 3926
       |||:::::::::   ::::::   ||||||:::   ::::::::
 613 ......SerValAspThrCysGluAspMetAspValThrGluAspMetGl 627

3927 CGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCC 3976
     :|||:::                    |||          :::::::::|
 627 uAspIle.....................ValLeuMetAlaAspLysSerH 637

3977 ATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCG 4026
     |||||                     |||::::::   |||:::|||
 637 isSer.....................TyrMetSerProGluMetAlaArg 646

4027 TCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGC 4076
        |||
 647 TrpAla.......................................... 648

4077 TGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCT 4126
                          :::||||||   |||
 649 .....................AspValLysTyrGly.............. 653

4127 TTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGC 4176
                                                 ::::::
 654 ..............................................AsnAsn 655

4177 AAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAG 4226
     ||||||   |||   |||      ||||||   :::   |||
 656 LysGlyAlaLeuValGluTyrLysValGlyThrSerMetThr........ 669

4227 GGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCG 4276

 669 ..................................................... 669

4277 ACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATT 4326
                                                     :::
 670 ..............................................Leu 670

4327 CCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGA 4376
     |||   :::|||   :::|||              :::||||||
 671 Pro...AlaThrTrpAlaGlu..............LysValLysAlaVa 681

4377 AGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAAC 4426
      :::||||||:::   :::          |||
 681 lLeuProLeuSerGlyIleCysValArgLysPro.............. 692
```

```
4427 CCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAG 4476

 692 ..................................................... 692

4477 GCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAA 4526
                                        |||   ::::::|||
 693 ..................................GlnPheSerLysPro.. 697

4527 ACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCA 4576

 697 ..................................................... 697

4577 TGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCG 4626
                                     ||||||     ::::::
 698 ..............................LeuAspGluGluAspAsp... 703

4627 ACTATCGCCACCAGGTTCTCAAACGCA...TTTACCTTTGTCGATAACTT 4673
                  |||:::||||||      |||      |||::::::||
 704 .........LeuArgLeuSerAsnMetAsnPhePheLysValSerAspLe 717

4674 GAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCT 4723
     ||||
 717 uLys.............................................. 718

4724 ATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGAT 4773
                                                  |||:::
 719 .....................................LeuLysLys 721

4774 ATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCA 4823
        :::|||   :::   |||   :::::::::   :::      :::::
 722 ThrIleThrProValValTyrThrGlyThrIleArgGluArgGlnMetLy 738

4824 GAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGT 4873
     ::::|||
 738 sAsnTyrIle.......................................... 741

4874 GCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCA 4923

 741 ..................................................... 741

4924 AACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGT 4973

 741 ..................................................... 741

4974 AGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACT 5023
               ||||||:::            |||             :::|
 742 ..........AspTyrLeuSerAlaSerLeuGlySerThrLeuGlyAsnL 755

5024 TGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGG 5073
     ||:::   :::
 755 euGluArgIle..................................... 758

5074 GCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGG 5123

 758 ..................................................... 758

5124 CGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCC 5173

 758 ..................................................... 758

5174 AAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCT 5223
                                                  :::   |||
 759 .....................................ValArgSer 761

5224 GAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGG.. 5271
     :::   :::|||   |||   ::::::   :::|||:::::::::::::
 762 AspTrpAsnGlyThrGlu...GluSerMetGlnThrPheGlyLeuTyrAs 777

5272 .TGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATC 5320
      ||||||             ||||||:::||||||
 777 pCysGlu..............LysCysLysTrpLeu.............. 784

5321 CAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTT 5370
```

```
                 |||   |||                 ::::::   ||||||
 785 ..........LeuLeuProAlaGluLysLysHisAlaTrpAlaValVal 797

5371 AAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGC 5420
        ::::::   :::        ::::::   :::|||:::|||
 798 LeuAlaSerAspAspThrThrArgIleIle...PheLeuSerTyrAspGl 813

5421 TATAGGCATT............GAGAGGACATACAGGGCGTTGCTTATGG 5458
        |||               ::::::::::::::   :::   ::::
 813 uSerGlySerProIleIleAspLysLysAsnTrpLysArgPheAlaValC 830

5459 CTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT 5508
     ::   :::      |||   ::::::         |||:::|||   |||
 830 ysSerGluThrLysValTyrSerValIleArgSerLeuGluValLeuAsn 846

5509 CAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGA 5558
     :::              ||||||   |||:::
 847 LysGluAlaIleValAspProGlyValHisIle..................  857

5559 TCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGT 5608
                 :::      ||||||:::
 858 .............ThrLeuValAspGlyVal...................  863

5609 ACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGAC 5658

 863 ..................................................  863

5659 GCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGT 5708

 863 ..................................................  863

5709 TTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCG 5758

 863 ..................................................  863

5759 GCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCT 5808

 863 ..................................................  863

5809 CCCCCAGGTGGCGGTAAGACGACGACGTTAGTG.................. 5841
        ||||||   ||||||||||:::   ::::::
 864 ...ProGlyCysGlyLysThrAlaGluIleIleAlaArgValAsnTrpLy 879

5842 .GACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATG 5890
        :::::::|||   :::|||:::      |||:::
 879 sThrAspLeuValLeuThrProGlyArgGluAlaAla..............  891

5891 TGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAAT 5940
                                    ::::::::::|||
 892 ......................AlaMetIleArgArgArgAlaCysAla  900

5941 TTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTT 5990
     |||:::             ||||||:::   |||   |||:::   :::
 901 LeuHisLysSerProValAlaThrAsnAspAsnValArgThrPheAspSe 917

5991 TATCGTC......AGGGGAATGTATAAA......AGGGTTTTGGTGGATG 6028
      :::|||      |||   ::::::|||         |||   |||||||
 917 rPheValMetAsnArgLysIlePheLysPheAspAlaValTyrValAspG 934

6029 AGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACC 6078
     ||      |||:::|||   |||||||||:::::::::::      :::
 934 luGlyLeuMetValHisThrGlyLeuLeuAsnPheAlaLeuLysIleSer 950

6079 GGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCAT 6128
     |||:::::::::::::::   |||||||||   :::|||||||||||||
 951 GlyCysLysLysAlaPheValPheGlyAspAlaLysGlnIleProPheIl 967

6129 AAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGA 6178
     |||||||      |||:::::::::|||               ||||||:
 967 eAsnArg......ValMetAsnPheAsp............TyrProLysG 978

6179 AG.................GAAAGCGTTGTATACACTTCTAAATCGTAC 6210
     ::                 ::::::|||               ::::::
```

```
 978 luLeuArgThrLeuIleValAspAsnValGluArgArgTyrValThrHis 994

6211 AGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGG 6260
     |||||||||   ||||||      :::|||:::::::::
 995 ArgCysProArgAspValThrSerPheLeuAsnThrIle...........1007

6261 AACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAG 6310
                           |||::::::   :::      :::||||
1008 ...................TyrLysAlaAlaValAlaThrThrSerProV1018

6311 TAGTAAGATCGCTGTCCAAA.........................AGG 6333
     |||||:::|||::::::                            |||
1018 alValHisSerValLysAlaIleLysValSerGlyAlaGlyIleLeuArg1034

6334 CCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTT 6383
     |||                ::::::::::::|||::::::   :::   ::
1035 Pro..............GluLeuThrLysIleLysGlyLysIle...Il 1045

6384 GTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAA 6433
     :   :::||||||   :::|||:::::::::   :::|||
1045 eThrPheThrGlnSerAspLysGlnSerLeuIleLysSer..........1058

6434 AAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACA 6483
                :::   |||   ||||||||||||   ||||||:::|||
1059 .....GlyTyrAsnAspValAsnThrValHisGluIleGlnGlyGluThr1073

6484 TTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATT 6533
     |||:::::::::::::::   |||:::      :::      :::|||::
1074 PheGluGluThrAlaValValArgAlaThrProThrProIleGlyLeuIl1090

6534 CACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCT 6580
     :::::::   :::||||||:::|||||||:::|||:::|||||||||::::
1090 eAlaArgAspSerProHisValLeuValAlaLeuThrArgHisThrLysA1107

6581 CACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACA 6630
     :::::||||||   ::::::         :::   |||   |||   :::
1107 laMetValTyrTyrThrVal.........ValPheAspAlaValThrSer1120

6631 TATATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACAC 6680
        |||:::|||:::      ::::::|||:::::::::::::|||
1121 IleIleAlaAspVal......GluLysValAspGlnSerIleLeuThrMe1135

6681 GTTCGCC 6687
      ||||||
1135 tPheAla 1137
```

```
seq_name: pir2:JQ1896

seq_documentation_block:
hypothetical 191.1K protein - raspberry bushy dwarf virus
C;Species: raspberry bushy dwarf virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-2000
C;Accession: JQ1896
R;Ziegler, A.; Natsuaki, T.; Mayo, M.A.; Jolly, C.A.; Murant, A.F.
J. Gen. Virol. 73, 3213-3218, 1992
A;Title: The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus.
A;Reference number: JQ1896; MUID:93107856
A;Accession: JQ1896
A;Molecule type: mRNA
A;Residues: 1-1694 <ZIE>
A;Cross-references: GB:S51557; NID:g262149; PIDN:AAB24599.1; PID:g262150
C;Genetics:
A;Map position: segment RNA-1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: RNA replication


alignment_scores:
            Quality:  291.00              Length:    477
              Ratio:  1.282                 Gaps:     15
 Percent Similarity:  47.589   Percent Identity: 25.577

alignment_block:
```

```
US-09-301-906-3 x JQ1896   ..

Align seg 1/1  to: JQ1896  from: 1  to: 1694

   5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
        |||   |||   ||||||:::   ||||||:::   :::   ::::::::
    725 ProGluThrProGluValThrMetThrGlyAsnProAsnAlaValIleAl 741

   5400 GGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGT 5449
        :|||   :::::::|||           :::::::|||      :::
    741 aGluAlaIleSerTyrCysArgAlaGluValAspArgIleGlyLys.... 756

   5450 TGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAA 5499
                                       :::|||   :::|||
    757 ............................LysCysGluArgIleLeu... 762

   5500 GTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTC 5549
               ::::::|||::::::      :::     |||   ::
    763 .........HisAlaTyrGlnAlaThrGlyAsnCysGlyGlyTyrLeuAs 776

   5550 GGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAG 5599
        :   ||||||   :::
    776 nAspThrAspAsnVal................................. 781

   5600 ACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTT............ 5637
                                     |||:::|||
    782 ......'.....................GlyValPheAspLysMetThr 788

   5638 .........ACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGA 5678
                    |||   |||:::   |||         :::
    789 SerTrpValGlnLysProLysGluPheAspHisGluPheGlyTrpAspGl 805

   5679 CAATCTGTTCGTA..............................TGCG 5695
         :::   |||:::
    805 ySerSerPheIleLysLeuSerTrpPheGlyLysIleProAspPheValG 822

   5696 ACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAA 5745
              ||||||      |||:::   ::::::         ||||||
    822 lyArgTyrLeuValValThrAspGlyThrArgValThrThrAsnLeuLys 838

   5746 GTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAA 5795
          :::|||         :::     :::     :::          ||
    839 PheSerArgGlnTyrAlaThrIleProAlaThrValThrProThrIleLy 855

   5796 GTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACG 5845
        |      ::::::      |||   ||||||||||||   :::|||   :
    855 sLeuValAspGlyValThrGlyCysGlyLysThrThrGluIleValArgA 872

   5846 AATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA 5895
        :::::      :::            ::::::      ||||||||
    872 rgTyrArgProGlyIleLeuIleLeuSerValCysLysAlaAsnVal... 887

   5896 AGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGA 5945
             ::::::|||:::      :::         |||   ::::::
    888 ......AspGluIleArgArgLysLeuAlaAlaValAspSerLysPhe.. 901

   5946 AGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCG 5995
                          ||||||:::|||   :::::::::      |
    902 .............IleArgThrValAspSerTyrLeuLeuSerProServ 914

   5996 TCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCAT 6045
        ||   |||         ::::::::::::||||||      :::   |||
    914 alThrGlySerCysAspGluLeuPheIleAspGluTyrGlyLeuSerHis 930

   6046 CAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCT 6095
           |||:::|||   |||::::::      :::|||      :::
    931 ProGlyIleLeuLeuLeuAlaIleHisIleSerGlyIleArgLysValTh 947

   6096 CTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGT 6145
         :::||||||||||   :::|||||||||||||   ||||||
    947 rLeuPheGlyAspSerGluGlnIleProPheCysAsnArg.......... 960

   6146 TTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTT..... 6189
```

```
                        |||   |||      |||   :::|||||||
    961 ...........LeuAlaAspPheProLeuLysTyrAsnSerValGluAsp 973

   6190 .....................GTATACACTTCTAAATCGTACAGGTGTCC 6218
                                        |||   :::||||||||||||
    974 ValGlyLeuAsnPheAspArgGluIleArgSerThrThrTyrArgCysPr 990

   6219 GTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAA 6268
        |   |||:::         |||:::::|||                  |
    990 oGlnAspIleThrLeuSerLeuGlnLysMet..............TyrL 1002

   6269 AGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGA 6318
        ||      |||   |||:::||||||                |||
   1002 ysThrLysProIleLysThrValSer............ThrValGluSer 1014

   6319 TCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAA 6368
        |||:::::::   :::||||||   :::      ::::::         ||
   1015 SerIleThrIleLysProIleLysSerGluPheGluIleProLeuProAs 1031

   6369 CGCT............GACGTGTACTTGTGCATGACCCAGTTGGAGAAGT 6406
        ||||                ::::|||:::||||||||||:::   ::::::|
   1031 nAlaPheAspGlyProValLeuTyrIleCysMetThrLysHisAspGluS 1048

   6407 CGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATG 6456
        ||   :::|||              ::::::   ::::::   |||
   1048 erLeuLeuLysLeuArgTrpAlaLysGluAsnIleSerSerGluValArg 1064

   6457 ACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAG 6506
        ||||||||   ||||||||   ::::::::::::||||||   |||||
   1065 ThrValHisAlaAlaGlnGlyLeuSerTyrLysAsnValValTyrPheAr 1081

   6507 GACGAAGAAAGCCGATGACTCCCTATTCACTAAACAA......CCGCATA 6550
        |      ::::::|||:::::::|||:::||||||:::      |||:::
   1081 gLeuThrArgThrAspAsnAspLeuTyrThrLysArgLysLeuProTyrH 1098

   6551 TACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTG 6600
          ||||||:::::|||||||||||   ::::::||||||:::::
   1098 isLeuValAlaIleSerArgHisThrAspLysIleValTyrCysThr... 1113

   6601 AGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACC 6650
                                       |||      :::|||:::|||
   1114 ........................ThrLysProGluAspSerSerAs 1121

   6651 TCAATCAGTATCCGACGCTTTGCTTCACACG 6681
           |||:::|||   ||||||   :::|||
   1121 pPheSerLeuSer...AlaLeuLysAsnThr 1130
```

```
seq_name: pir1:P1BVBB

seq_documentation_block:
RNA 1a protein - broad bean mottle virus (strain Bawden)
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase
C;Species: broad bean mottle virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 05-May-2000
C;Accession: A41699
R;Dzianott, A.M.; Bujarski, J.J.
Virology 185, 553-562, 1991
A;Title: The nucleotide sequence and genome organization of the RNA-1 segment in two
A;Reference number: A41699; MUID:92074218
A;Accession: A41699
A;Molecule type: genomic RNA
A;Residues: 1-966 <DZI>
A;Cross-references: GB:M65138; NID:g210662; PIDN:AAA42740.1; PID:g210663
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine
F;80-168/Domain: methyltransferase #status predicted <MTF>
F;688-944/Domain: helicase #status predicted <HHG>
F;690-697/Region: nucleotide-binding motif A (P-loop)


alignment_scores:
            Quality:  290.00              Length:    775
              Ratio:  0.773                 Gaps:     33
```

```
 Percent Similarity:  48.387   Percent Identity: 23.613

alignment_block:
 US-09-301-906-3 x P1BVBB   ..

 Align seg 1/1  to: P1BVBB  from: 1  to: 966

    4580 TTAGAACGTGTTCCTGCGGTGTGCAGCTGGATG...........TGTAC 4617
         |||||||||:::         ||||||   |||              ::
     275 LeuGluArgMetLeuLeuLysCysSerIleMetThrTyrLysIleIleAl 291

    4618 AACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTT.ACCTTTGTCG 4666
         :|||:::         ||||||   |||      |||      :::   :
     291 aThrAsnValArgCysProProGluSerLeuArgHisCysIleTrpPheG 308

    4667 ATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGG 4716
         ::|||:::::         ||||||        ::::::      |||
     308 luAsnIleSerGlnTyrLeuAlaVal.........GlnIleProIleGly 321

    4717 TATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCT 4766
         |||:::   |||                        |||      :::
     322 TyrAsnLeuAsn....................AspTrpLysThrValAr 331

    4767 AGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTT 4816
                  |||:::   :::      ::::::      :::|||:
     331 gValAlaArgAlaThrValArgGluValGluGluIleSerPheArgCysP 348

    4817 TAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGAC 4866
         ::   ::: :::      |||:::|||      :::|||:::
     348 heLysGlu.AsnLysAspTrpThrGluAsnMetArgSerValAlaSerIl 364

    4867 GAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGG 4916
             ||||||      :::   :::::: ::::::||||||||
     364 eLeuSerAlaLysSerSerThrValIle.IleAsnGlyGlnSerIleMet 380

    4917 GAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCA 4966
         :::|||:::   ||||||   ::::::
     381 SerGlyGluArgLeuAspValLeuGlu..................... 389

    4967 TAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGG 5016
                                    :::   :::   ||||||
     390 .......................TyrHisLeuValAlaPheSerLeuT 398

    5017 ACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGAC 5066
                  :::::::|||   ::::::   |||      |||   :::
     398 hrLeuAsnLeuTyrGlnLysTyrGluLysLeuArgAsnPheGlnGlyGlu 414

    5067 GTTCAGGGCAAC.................TCGGCGCGTCTT...TGGTG 5095
         ::::::   :::                 ::::::||||||   |||||
     415 LeuGluTrpLysGlyTrpAlaAsnHisPheLysThrArgLeuTrpTrpCy 431

    5096 TAGGCAGGATGTTGCAGTTAGCCGGCGGC..................... 5124
         | ||||||   ::::::      ||||||
     431 s.GlyArgThrValSerThrGluGlyGlyPheLeuArgAsnPheLeuAla 447

    5124 ................................................. 5124

     448 AspLysIleProTrpLeuLysLeuAsnThrTyrAlaAspSerLeuAspPh 464

    5125 ..........GTGTCGGATGAGAAGTCA.........CCAGGTGTTCCAA 5155
                  :::|||:::   :::|||           :::||||||:
     464 eIleThrLysIleSerGluValGluSerPheGluValAspSerValProT 481

    5156 ACCAGCAACCACAGAGC...CAAGGTGCTACCAGAACAATCACACCAAAA 5202
         ::::::::   :::|||               ::::::|||:::   :::
     481 hrSerArgLeuArgSerPhePheGlnLysGluGluAsnIleValGluArg 497

    5203 TCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAG 5252
         ::::::::::::   :::|||   ::::::   |||   :::|||
     498 AlaAlaSerGluIleMetSerAlaAsnAlaArgArgIleAlaLys..... 512

    5253 GTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGT 5302
                                                  |||:::|||

     513 .....................................LysAlaGluM 516

    5303 GGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCA 5352
              :::      :::      :::           :::::::::|||
     516 etSerLysGluPheAspAspPheValAspAlaProGluGluPheAlaPro 532

    5353 ......ATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGT 5396
                 ::::::   ||||||::::::|||   :::::::::|||:::
     533 GluAspValValGluGluValIleAsnThrProValThrGlnAspValLy 549

    5397 CGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGG 5446
          ::::::      |||                        :::|||   |||:
     549 sLeuArgGlnSerLys..................ProGluThrAlaArgS 560

    5447 CGTTGCTTATG..............GCTAGAAATATTGCCGTCAC.... 5477
         ::::::::::                  :::|||   |||:::
     560 erIleValLeuAspProAspAlaValLeuLysAsnGlyAlaIleAsnGlu 576

    5478 .........TACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATG 5518
                  |||   :::|||      |||:::|||   |||::::|| :
     577 PheAlaAspTyrSerLysArgLeuHisGluAsnThrValSerAsnLeu.A 593

    5519 AATCACTA..................CCGGGCTTTCACGTTTACAAGTCG 5550
         ::   |||                     |||   ::::::::::::::
     593 rgHisLeuTrpThrLeuMetGlyCysArgGlyAsnGluIleHisAsnLys 609

    5551 GGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGA.. 5598
         ::::::      :::|||   ::::::|||   ::::::|||:::
     610 SerValAlaGluThrTyrHisArgValAspAspMetValAsnValHisPh 626

    5599 ......................GACCTACCGTACGTACTCATAGCTGAAA 5626
                                    |||   |||          :::
     626 eProAsnGlyHisTrpMetTyrProLeuLysTyrGluTyrThrValGlyT 643

    5627 AAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGC 5676
           :::            |||:::      :::      :::   :::
     643 yrAsnAspGlyGlyLeuGlyGluLysPheGluAsnGluLeuTyrValVal 659

    5677 GACAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAA 5726
         |||:::          |||   :::      ::::::   ||||||   ::
     660 AspLysThrCysSerCysAlaAsnAlaLysAlaIleAlaAspAlaCysLy 676

    5727 TTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGT 5776
         :   ::::::|||                |||:::::::|||
     676 sLysValSerAla.........ProThrCysSerValVal.......... 686

    5777 TTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAG 5826
                                      ::::::      |||   ||||||
     687 ....................MetValAspGlyValAlaGlyCysGlyLys 696

    5827 ACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCAC 5876
         ||||||::::::   :::   |||            ::::::
     697 ThrThrAlaIleLysGluThrPhe......ArgPheGluLysAspIleIl 711

    5877 CATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGA 5926
          :::|||||||||      |||:::||||||::::::   |||:::
     711 eValThrAlaAsnArgLysSerAlaGluAspValArgLysAlaIle.... 726

    5927 AGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCC 5976
                           |||      |||   :::   |||
     727 .................PheGlyAspAlaSerAspSerGluValAlaLeu 737

    5977 AGGGTGGTA............AACTTTATCGTCAGGGGAATGTATAAA.. 6012
         :::||||||             :::   |||::::::|||:::   :::
     738 LysValValArgThrAlaAspSerAlaIleMetHisGlyLeuProGluCy 754

    6013 ....AGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTAC 6058
             |||:::||||||||||||:::   :::::|||   |||   |||
     754 sHisArgLeuLeuValAspGluAlaGlyLeuLeuHisTyrGlyGlnLeuL 771

    6059 AACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGAC 6108
              :::            :::||||||   |||   |||||||||
     771 euAlaValAlaAspLeuCysLysCysSerGluValLeuAlaPheGlyAsp 787
```

```
 6109 ATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGAT.. 6156
       :::||||||   |||   :::|||:::   :::||||||||||
  788 ThrGluGlnIleSerPheLysSerArgAspAlaThrPheArgMetLysTy 804

 6157 .TGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAAT 6205
       |||   :::         |||::::::   |||         ||||||:
  804 rCysAsnIleGluTyrAspLysArgAspIleVal.........SerLysT 818

 6206 CGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTA 6255
      :::::|||||||||   ||||||         :::|||:::::   :::
  818 hrPheArgCysProGlnAspVal.........ValSerAlaValLysIle 831

 6256 AGGGGAACGGAAAAGTGT.................TACCCTGAAAAGGT 6287
             :::||||||                |||        ||
  832 Leu...LysArgLysCysAlaAsnArgSerSerLysTyrAsnGlyTrpVa 847

 6288 CGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAA 6337
      |   |||:::|||         |||   :::||||||||||||     |
  847 lSerSerSerLys.........ValGluLysSerLeuSerLysSerArgI 861

 6338 TTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGC 6387
      ||   :::   ::::::|||:::      :::         ||||||
  861 leValSerIleAsnGlnValSerMetGluLysHisLysPheTyrLeuThr 877

 6388 ATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGG 6437
      ||||||:::   :::|||:::   :::            |||      ||
  878 MetThrGluAlaAspLysAlaAlaLeuCysSerArgAlaLysAspValGl 894

 6438 A..............AAAGAAACACCAGTGATGACAGTGCATGAAGCAC 6472
      |                |||:::   :::   ||||||||||||||||
  894 yLeuAspLysThrTrpValGluSerAsnMetGluThrValHisGluAlaG 911

 6473 AGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGAT 6522
      ||||||||:::   :::   ||||||||   |||   |||::::::
  911 lnGlyLysAlaValAspHisValValLeuValArgLeuLysSerThrLys 927

 6523 GACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACA 6572
        :::||||||   ::::::   :::   ||||||:::|||:::||||
  928 CysAspLeuPheLysSerGluGluTyrCysLeuValAlaLeuThrArgHi 944

 6573 CACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGG 6622
      |   |||::::::   |||         ::::::         |||   :
  944 sLysArgThrPheGluTyrLeuTyrAsnGlyAspLeuGlyGlyAspLeuI 961

 6623 TCGGCACATATATTAGC 6639
      :::::   |||::::::
  961 leSerPheTyrValThr 966

seq_name: pir2:JQ2183

seq_documentation_block:
hypothetical 216.5K protein - apple chlorotic leaf spot virus (isolate P-205)
N;Alternate names: ORF 1 protein
C;Species: apple chlorotic leaf spot virus
C;Date: 30-Sep-1993 #sequence_revision 20-Aug-1994 #text_change 20-Jun-2000
C;Accession: JQ2183
R;Sato, K.; Yoshikawa, N.; Takahashi, T.
J. Gen. Virol. 74, 1927-1931, 1993
A;Title: Complete nucleotide sequence of the genome of an apple isolate of apple chlorot
A;Reference number: JQ2183; MUID:93389448
A;Accession: JQ2183
A;Molecule type: mRNA
A;Residues: 1-1885 <SAT>
A;Cross-references: GB:D14996; NID:g434059; PIDN:BAA03641.1; PID:g453240
C;Superfamily: eggplant mosaic virus RNA-directed RNA polymerase
C;Keywords: ATP


alignment_scores:
            Quality:  281.00            Length:   1401
              Ratio:  0.466              Gaps:     74
 Percent Similarity:  43.041   Percent Identity: 20.271
```

```
alignment_block:
 US-09-301-906-3 x JQ2183   ..

 Align seg 1/1  to: JQ2183  from: 1  to: 1885

 2938 TCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCAT 2987
      |||:::::::::::::::::|||:::   ||||||           |||::
  235 SerSerSerIleLeuValArgAsnLysAlaThr...........GlyVa 247

 2988 TACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTA.... 3033
      :         :::::::::   ||||||   |||:::   |||
  247 lGluIleArgTyrGlnVal...SerLeuValTyrSerLeuGlySerHisH 263

 3034 ...................ACTGAAGATGAAAGGCTTGATGCTCTGGTG 3063
                           |||||||||               |||:::
  263 isLeuPheHisIleTyrProThrGluAsp..............LeuMet 274

 3064 CGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGTCGCG 3113
      :::   |||            :::|||      ::::::::::::::
  275 LysGluGluValArgArgPheGlyProTyrAspLeuPheAspValGlySe 291

 3114 CGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAG 3163
       :::   |||:::   |||                           :::|
  291 rLeuPheValLysProVal.........................ArgV 299

 3164 TTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATC 3213
      ||   ::::::||||||         |||:::::::::         |||
  299 alProIleGlnAspPheProLeuSerValPheLysLys.........Ile 312

 3214 GGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAG 3263
         |||::::::::::||||||          :::
  313 PheIleTyrLeuSerSerLeuLysLysProAspVal.............. 324

 3264 GCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAA 3313

  324 ..................................................... 324

 3314 CTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACC 3363
                      ::::::|||::::::   ||||||   |||:::
  325 .................GlnSerAlaValAlaLysLeuArgGlnLeuSer 335

 3364 TCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTC 3413
      ::::::   :::|||      |||:::   :::   :::   |||:::||
  336 AspAlaAspIleSerIleGluSerValPheMetValGlnGluPheAlaSe 352

 3414 GGCGGTGACCGTCACTAGGGCG.......................... 3435
      |   |||      :::   :::
  352 rArgValGluLysAsnGlyValGlySerTrpSerCysSerPheTrpGluC 369

 3436 ......................ACCGTAGCTAAACGTCAAGTCCCCTTA 3462
                                  :::      |||:::|||   |||
  369 ysMetLysAspTrpPhePheAspLysLeuProTyrArgGluVal...Leu 384

 3463 GCGTTGCTATCGTTTTCTACCTCATACGCC......ATTTCCGGCTGCAG 3506
            ::::::::::::::::::::::::      :::         ::
  385 GluLysIleGlyLeuAlaAsnAspPheThrArgArgLeuMetLysIleLy 401

 3507 TATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCT 3556
      :   |||::::::      |||:::      |||   |||         :
  401 sProLeuAlaPheAspIleHisThrThrAspArgProLeuThrValArgM 418

 3557 TTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAG 3606
      ::   :::      :::   |||   |||   |||:::
  418 etValIleAspGlnIleTrpGlyGluArgGlnSerSer............ 430

 3607 TTTGGAGGCTTCTCCAATAATTGGTGC...GCTGTTCCCGAGGTTGTTTG 3653
                              |||      ||||||::::::::::
  431 ......................CysAspAspValProAsnIleIlePh 439

 3654 GCGAGGGAAGAGT.........GTCAGCTCATTGTTACTGCCTATTACGC 3694
      :   |||::::::               :::||||||      :::
  439 eTyrGlyArgLysGluTrpLeuGluHisGlyLeuLeuProLysValLysL 456
```

```
3695 TAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCTCAA 3744
      |||:::::::   ::::::   |||      ::::::|||
 456 ysGlyLeuAlaLysLeuValProGly......ArgGluThr..... 467

3745 CTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAGACGCT 3794
                           |||   :::   |||   |||   ::
 468 .....................GlyGlySerAspTyrProGluGluIl 476

3795 TAGGTATTACCGGGACTTT.....................GACTATG 3820
      :     |||   |||:::                      :::||||
 476 e......TyrSerAspLeuLeuSerSerThrSerIleTrpArgSerTyrA 491

3821 ACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGAC 3870
     |||||:: |||   |||                |||      |||:::::
 491 spGluAsn.LeuArgHis...........ArgLysAlaSerProIleVa 503

3871 GATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAA 3920
     :|||      :::          |||:::|||      :::|||:::
 503 lIleLeuLysSerGluLysAlaTyrSerGluAlaProGlyPheSerSerA 520

3921 TGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAG 3970
       |||   :::        :::|||   |||:::   ::::::::::::
 520 snCysIleSerLeuCysSerThrProPheGlyGlu...ValIleGluArg 535

3971 AGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTT 4020
       |||::::::::|||                    |||:::|||
 536 ThrProPheGluValGlu.................ArgGluArgLy 545

4021 GCCCCGT...CTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTC 4067
        |||   |||::::::         |||      ||||||::::::
 545 sLysArgGluLeuSerPheGlyCysLeuAspPheHisIleArgLysMetL 562

4068 TGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCG 4117
       |||                                |||   :::|||
 562 ysVal..............................LysAspAlaSer 567

4118 ACGATGTCTTTCACCAACAGTCTAGTGGGGAAA..................  4150
       :::|||:::      :::      :::|||
 568 GluLeuSerIleLysLeuAspGluGlnAsnLysGluGlyLeuArgArgGl 584

4151 ........................CGGCTCGTGAGGTTGAGGTGGA 4172
                              |||::::::   ::::::
 584 nLysArgLysGluLysAlaLysLysThrArgMetIleProValHisLeuL 601

4173 CGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAG 4222
           |||   :::|||||||   ::::::::: |||   ::::::
 601 euGluLeuGlySerAspGlnLysGluLysAsnLeu.GlyGlnGluAlaSe 617

4223 AAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAA 4272
      |||||||||:::   ::                          |||::::
 617 rLysGlyLysGlyIleGlu.....................GlnGluG 626

4273 GGCGACAGGCAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGA 4322
     ::||| |||:::      ::::::::::::   |||:::|||
 626 luArg.ArgLysSerAspGluAlaLysPheAspSerGlyPro........ 639

4323 TATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAAC 4372

 639 ..................................................  639

4373 AAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAG 4422
                           ||||||:::      |||      |||
 640 .....................SerGlyValCysSerIleLysAlaGlu 648

4423 AAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGG 4472
     :::|||:::   :::
 649 AsnProValAspAla..................................  653

4473 CAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGC 4522
                  |||:::||||||
 654 ..........GlnHisIleAlaGluProValPro.............. 661
```

```
4523 CCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTT 4572
                    |||   :::|||:::      ||||||:::
 662 ...........CysLeuLysLeuAsnAsp...LeuIleGlyLysGlu... 672

4573 TGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGA 4622
           :::   ||||||   |||:::         |||   |||::
 673 .........LysIleCysSerSerGlyLeuIleLysThrValGlyAsnAs 686

4623 A...GCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATA 4669
     :      |||:::|||   ::::::::::::           :::   :
 686 pTyrLeuThrLeuAlaArgGlnIleGluAsp......MetProLeuAlaG 701

4670 ACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTAT 4719
     ::|||||:::|||:::|||::::::|||      :::
 701 lnLeuLysAsnArgLysAlaAlaTyrPheCys...IleAspTyrProMet 716

4720 ACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCT......CGTGC 4763
     :::|||            |||   :::   ::::::|||      :::||
 717 ValTyr.........PheHisAspLysIleSerTyrProThrPheGluAl 730

4764 CCTAGAGGATATCTTAACGGCAATT..............AAGTACCCAA 4798
     |      :::|||         |||               |||:::   :
 730 aThrGlyGluIleLysGlnIleIleMetArgAlaArgAspLysTrpGlyA 747

4799 GCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTA 4848
     ::   |||:::   :::|||:::|||   |||:::   |||      :::
 747 laAsnPheAsnSerAlaLeuIleGlnValTyrAsnAspGlyCysArgLeu 763

4849 CCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTT 4898
     |||:::|||:::|||:::||||||||||||   :::|||:::   |||||
 764 ProLeuHisSerAspAsnGluGluCysTyr...AspAspAspGluIleLe 779

4899 GACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGG 4948
     ||||:::|||:::||||||   |||:::|||   |||   |||:::
 779 uThrIleAsnValValGlyAspAlaLysPheHisThrThrCysHis.... 794

4949 GTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATG 4998
       |||:::      :::|||::::::   :::||||||   :::||||||
 795 ..GlyGlu......IleIleAspLeuArgGlnGlyAspGluIleLeuMet 808

4999 CCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGA 5048
     |||   |||:::|||:::   :::   |||:::|||:::      :::||
 809 ProGlyGlyTyrGlnLysMetAsnLysHisAlaValGluValAlaSerGl 825

5049 AGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAG 5098
     |||||||   |||:::|||:::|||:::   :::|||   |||:::
 825 uGlyArgThrSerValThrLeuArgValHisLysArgAspPheSerPheG 842

5099 GCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGAT.............. 5133
            |||::::::      |||      |||
 842 luSerLysLeuArgPheIleLysGlyLysPheAspCysLeuPheValSer 858

5134 ..................GAGAAGTCA.....................CC 5144
                        :::|||                      ||
 859 IleAlaGluIleIleHisLysLysProGluGluValMetMetPheIlePr 875

5145 AGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACC.......... 5184
     |   :::   ::::::      ::::::::::|||::::::
 875 oHisIleValAspArgCysValSerAsnArgGlyCysSerLeuAspAspA 892

5185 ..AGAACAATCACACCAAAA.............................. 5202
       |||:::|||      |||
 892 laArgAlaIleCysGluLysTyrGluIleLysIleGluCysGluGlyAsp 908

5203 ................................TCGGGGGGCAAGGCTCT 5219
                                     |||   |||:::   ||
 909 CysGlyLeuValGluCysGlyThrIleGlyLeuSerValGlyArgMetLe 925

5220 ATCTGAGGGA....................................... 5229
     |   :::|||
 925 uLeuArgGlyAsnHisPheThrValAlaSerValArgArgSerSerMetA 942

5230 ..............AGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCG 5265
```

```
                    |||:::::::|||:::|||:::::::::::::
  942 spSerLeuAlaAsnSerSerLysGluLeuLysSerAsnGlyValLeu... 957

 5266 ATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGA 5315
                     |||:::|||                 ::::::
  958 ..............AspHisVal............MetPheAsnPheHi 965

 5316 TAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAG 5365
       :::   :::   ::::::|||||||||         :::|||   |||
  965 sLysArgLeuArgSerValGluProAsp.........LeuThrValGluG 979

 5366 TGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTAT 5415
       :::|||:::   :::|||:::   :::       :::   ||||||
  979 luIleLysValAspSerSerArgAlaGly......LysLeuLeuLysSer 993

 5416 CTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAA 5465
      |||      |||:::
  994 LeuMetAspGlyMet..................................... 998

 5466 TATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTT 5515

  998 ..................................................... 998

 5516 ATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATT 5565
                                         |||   ::::::
  999 .....................................ThrGlyIleVal 1002

 5566 TTTCAT...TCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGT 5612
         |||   ||||||::::::|||   |||:::
 1003 SerHisAsnSerThrHisGluGlyTrpArgMet.................. 1013

 5613 ACTCATAGCTGAAAAAGGTATC............................. 5634
                     |||||||||
 1014 ..........IleLysGlyIleAsnSerThrSerGluMetArgSerPheM 1027

 5635 .....TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGAC 5679
           :::   |||||||||:::   :::         :::   ||||||
 1027 etSerMetIleLysGlyLysSerGluGlu.........ThrArgGlyAsp 1040

 5680 AATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTT 5729
         ||||||                                |||   ::::::
 1041 ...LeuPhe..........................AspArgValGlnGl 1047

 5730 GATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTA 5779
       ::::::   :::||||||          |||:::   |||   :::
 1047 uLeuAsnPheMetLysValLysIleTyrGlyIlePheGlyPheAla.... 1062

 5780 AGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACG 5829
                                        |||:::||||||:::
 1063 ................................GlySerGlyLysSer 1067

 5830 ACGACGTTAGTGGAC............GAATTCGTTAAGTCACCCAATAG 5867
         ::::::   :::            ||||||      |||   :::
 1068 HisAlaIleGlnAsnLeuIleGlnThrGluPheLysGlySerGlnGlyVa 1084

 5868 CACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGG 5917
         :::          :::      :::::::::      :::   :
 1084 lMetValIleCysProArgArgPheLeuAlaLysAspTrpSerGluLysG 1101

 5918 CGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACA 5967
       ::|||   ::::::|||      ::::::   ::::::||||||
 1101 lyValAspGluLysAsp......IleLysThrPheGluSerAlaLeuLys 1115

 5968 GTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGT 6017
         :::                  |||:::|||      ||||||::
 1116 SerAsp...................ValLysGly......LysArgLe 1123

 6018 T...TTGGTGGATGAGGTGCACATGATGCATCAAGGC.........TTAC 6055
      :   ::::::||||||:::   ::::::   :::|||         ||||
 1123 uPheIleLeuAspGluIleSerLeuLeuProLysGlyPheThrAspLeuL 1140

 6056 TACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGA 6105
      ||:::|||   :::         |||
```

```
 1140 euMetLeuLysMetHisMetGluGly..................... 1148

 6106 GACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGA 6155

 1148 ..................................................... 1148

 6156 TTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATAC......... 6195
                 :::   ||||||::::::::::|||
 1149 .............IleLeuLysLysSerThrIleValCysIleGlyAspP 1161

 6196 ..ACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
         :::   :::|||   ||||||   |||      ||||||
 1161 roLeuGlnAlaGlyTyrPheCysProLysAspAspAsnTyrLeuSerArg 1177

 6244 TCAATGACCGTAAGGGGA.................ACGGAAAAGTGTTA 6275
      :::      ::::::                    :::   |||   ||
 1178 GluGlyGluIleLysArgLeuPheLysAspGlyValAsnTyrLysTrpTy 1194

 6276 CCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGT 6325
      |      ::::::              :::|||   ::::::::::::|||:
 1194 rSerTyrArgIle.............AsnLysPheIleAlaLysLysLeuA 1207

 6326 CCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGAC 6375
      ::         |||   |||   :::|||         |||:::   :::
 1207 la.........IleGluThrMetAsnAspPheIleGlyIleAspGluGln 1220

 6376 GTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCG.. 6423
                        :::|||      :::||||||      :::
 1221 ...................AsnLeuIleTyrLysAspMetProSerAlaHi 1231

 6424 .......TTGAAGGGAAAAGGAAAAGAAACACCA................ 6450
             :::::::::|||||::::::
 1231 sHisPheMetGluAsnLysGlyAsnHisIleGluValIleLeuValProS 1248

 6451 .....................................GTGATGACAGTG 6462
                                           |||||||||
 1248 erMetIleGluLysGluLeuTyrSerAsnTyrGlyAsnValMetThrPhe 1264

 6463 CATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAA 6512
         |||:::||||||   ||||||:::      |||:::      :::::
 1265 GlyGluSerGlnGlyLeuThrPheAsnCysGlyValIleValLeuSerGl 1281

 6513 GAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTT 6562
      ::::|||            :::   ::::::   ||||||:::|||::::
 1281 uGluAlaLys.....LeuCysSerAspAlaHisIleMetValAlaI 1295

 6563 TGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAG 6609
      :::::|||      |||::::::   :::   ||||||:::||||||
 1295 leThrArgPheArgArgGlyPheCysPhe...AlaLeuGlySerLys 1309
```

This Page Blank (uspto)

```
OM of: US-09-301-906-3 to: SwissProt_39:*   out_format : pfs

Date: Jan 31, 2001  4:28 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09301906/runat_31012001_153038_9697/app_query.fasta_1.6881
 -DB=SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09301906_@CGN1_1_373 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-3
Query length: 6714
Database: SwissProt_39:*
Database sequences: 88757
Database length: 32294092
Search time (sec): 253.200000


score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
SwissProt_39:RRPO_SHVX +  414.00  445.71  1.9e-18   1718  ! Q04575 shallot virus x (shvx
SwissProt_39:RRPO_PMV +  404.00  435.47  7.9e-18   1547  ! P20951 papaya mosaic potexvir
SwissProt_39:194K_TRVSY +  338.50  359.55  1.2e-13   1707  ! P05080 tobacco rattle virus
SwissProt_39:RRPO_SHMV +  337.00  358.36  1.5e-13   1629  ! P89202 sunn-hemp mosaic viru
SwissProt_39:V125_AMVLE +  329.50  353.98  3.8e-13   1126  ! P03589 alfalfa mosaic virus
SwissProt_39:RRPO_TOMS1 +  313.00  331.04  5.0e-12   1616  ! Q9ykd6 tomato mosaic virus
SwissProt_39:RRPO_TOML +  309.00  326.48  8.9e-12   1616  ! P03587 tomato mosaic virus (
SwissProt_39:RRPO_TOMK2 +  303.00  319.62  2.1e-11   1616  ! P89676 tomato mosaic virus
SwissProt_39:RRPO_TOMK1 +  302.00  318.48  2.5e-11   1616  ! Q9q1t8 tomato mosaic virus
SwissProt_39:RRPO_CGMVS +  298.00  313.69  4.5e-11   1648  ! P19523 cucumber green mottl
SwissProt_39:RRPO_PMMVJ +  297.00  312.80  5.2e-11   1612  ! P89657 pepper mild mottle v
SwissProt_39:RRPO_PMMVS +  297.00  312.80  5.2e-11   1612  ! P29098 pepper mild mottle v
SwissProt_39:V1A_BBMV +  290.00  310.61  1.1e-10    966  ! Q00020 broad bean mottle virus
SwissProt_39:RRPO_TMVRA +  288.00  302.49  1.9e-10   1616  ! Q98745 tobacco mosaic virus
SwissProt_39:RRPO_TMVB +  285.00  299.06  3.0e-10   1616  ! Q9yq08 tobacco mosaic virus
SwissProt_39:RRPO_ACLSA +  281.00  292.75  5.8e-10   1885  ! P54891 apple chlorotic leaf
SwissProt_39:RRPO_ORSVS +  279.00  292.24  7.2e-10   1612  ! Q84133 odontoglossum ringsp
SwissProt_39:RRPO_ORSVC +  277.00  289.95  9.7e-10   1612  ! P89659 odontoglossum ringsp
SwissProt_39:V1A_PSVJ +  275.50  293.60  9.7e-10   1005  ! P28726 peanut stunt virus (st
SwissProt_39:V1A_BMV +  275.00  293.53  1.0e-09    961  ! P03588 brome mosaic virus (bmv)
SwissProt_39:RRPO_TMVOM +  275.00  287.64  1.3e-09   1616  ! O93058 tobacco mosaic virus
SwissProt_39:V1A_CMVFN +  274.00  292.02  1.2e-09    993  ! P17769 cucumber mosaic virus
SwissProt_39:VRNA_BSMV +  273.50  289.89  1.4e-09   1139  ! P17595 barley stripe mosaic
SwissProt_39:V1A_CMVNT +  273.00  290.88  1.4e-09    993  ! O40976 cucumber mosaic virus
SwissProt_39:V1A_CMVO +  273.00  290.88  1.4e-09    993  ! P20122 cucumber mosaic virus (
SwissProt_39:RRPO_TMGMV +  272.50  284.84  1.9e-09   1609  ! P18339 tobacco mild green m
SwissProt_39:FIG2_YEAST -  270.50  282.55  2.5e-09   1609  ! P25653 saccharomyces cerevi
SwissProt_39:RRPO_TMOB +  270.50  282.50  2.5e-09   1616  ! P90211 tobamovirus ob. rna-d
SwissProt_39:V1A_CMVII +  270.00  287.45  2.2e-09    993  ! Q83270 cucumber mosaic virus
SwissProt_39:V1A_CMVIX +  269.00  286.31  2.5e-09    993  ! Q66121 cucumber mosaic virus
SwissProt_39:RRPO_TMV +  267.00  278.51  4.2e-09   1616  ! P03586 tobacco mosaic virus (
SwissProt_39:RRPO_TMVKR +  264.00  275.08  6.5e-09   1616  ! P30738 tobacco mosaic virus
SwissProt_39:V1A_CMVQ +  263.00  279.48  6.0e-09    991  ! P06011 cucumber mosaic virus (
SwissProt_39:RRPO_ACLSP +  261.00  269.91  1.1e-08   1884  ! P27738 apple chlorotic leaf
SwissProt_39:V1A_CMVY +  253.50  268.61  2.4e-08    993  ! Q83264 cucumber mosaic virus (
SwissProt_39:V1A_CCMV +  252.00  267.30  3.0e-08    958  ! P27752 cowpea chlorotic mottle
SwissProt_39:YJ9P_YEAST -  241.50  253.13  1.5e-07   1161  ! P47179 saccharomyces cerevi
SwissProt_39:AMYH_YEAST -  240.00  249.56  2.0e-07   1367  ! P08640 saccharomyces cerevi
SwissProt_39:RRPO_TVCV +  238.00  245.49  2.9e-07   1601  ! Q88920 turnip vein-clearing
SwissProt_39:V1A_TAV +  237.00  249.76  2.7e-07    993  ! P28931 tomato aspermy virus (ta
SwissProt_39:YM96_YEAST -  224.50  233.92  1.8e-06   1140  ! Q04893 saccharomyces cer
SwissProt_39:RRPO_CRMV +  218.00  222.67  5.5e-06   1597  ! Q66220 chinese rape mosai
SwissProt_39:MUC2_HUMAN -  218.00  209.34  9.3e-06   5179  ! Q02817 homo sapiens (hum
SwissProt_39:AGA1_YEAST -  198.50  209.35  6.6e-05    725  ! P32323 saccharomyces cere
SwissProt_39:MSB2_YEAST -  195.50  199.26  0.0001    1306  ! P32334 saccharomyces cer

seq_name: SwissProt_39:RRPO_SHVX

seq_documentation_block:
ID   RRPO_SHVX      STANDARD;      PRT;  1718 AA.
AC   Q04575;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA REPLICATION PROTEIN (194 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
DE   DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
OS   Shallot virus X (ShVX).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93019008; PubMed=1339468;
RA   Kanyuka K.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,
RA   Ryabov E.V., Zavriev S.K.;
RT   "Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal
RT   cistron closely related to those of potexviruses and a unique
RT   arrangement of the 3'-proximal cistrons.";
RL   J. Gen. Virol. 73:2553-2560(1992).
CC   -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC       POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M97264; AAA47787.1; -.
DR   PIR; JQ1734; JQ1734.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW   Transferase.
FT   NP_BIND     915    922       ATP (POTENTIAL).
SQ   SEQUENCE   1718 AA;  194529 MW;  820FEAE1EB62415C CRC64;


alignment_scores:
           Quality:  414.00              Length:   1025
             Ratio:  0.828                 Gaps:     51
 Percent Similarity:  48.780   Percent Identity: 22.244

alignment_block:
 US-09-301-906-3 x RRPO_SHVX   ..

 Align seg 1/1  to: RRPO_SHVX  from: 1  to: 1718

    3974 CCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCC 4023
         |||:::|||:::             |||   :::   :::::::|||
     411 ProValHisPhe.............ThrThrGlnProLysArgArgGluAl 423

    4024 CCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTAC 4073
              :::||||||   ::::::
     423 aValGlyGluLeuTrpPheGlnGlu........................... 431

    4074 CGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATG 4123
                    |||                        :::|||||||||||||
     432 .....ProLysTrp.....................SerValSerThrMet 439

    4124 TCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGAC 4173
              |||:::                 :::   :::::::::   |||||
     440 ......ThrGlnProArgLysGluHisHisArgLeuGlnMetThrTrpTh 454
```

```
4174 G.....GCAAAGG 4181
     |          ::::::
 454 rLeuLeuAlaTrpPheHisGlnLeuGluSerSerGlySerMetSerGluP 471

4182 GTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGC.........CGACACAAG 4222
       |||::::::::::   |||:::   :::|||         ::::::
 471 roCysAsnAsnSerGluSerThrProGlnArgThrAlaThrSerGlnGln 487

4223 AAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAA 4272
     |||:::   :::|||           :::|||:::   |||:::   ::
 488 LysAlaAlaLysLeuThrThrSerGlnLysHisAsnArgArgThrAspGl 504

4273 GGCGACA...GGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGC 4319
     ::::|||   :::   ||||||   |||                    |
 504 nThrThrMetAsnProGlnTyrProPro.....................L 514

4320 TGATATTCCAAAGGTCACCC...AGTCCGAGGTACATGCTCAGAAAGAAG 4366
     ||::::::     :::|||     ||||||:::   |||   ||||||
 514 euMetLeuThrIleAlaProMetMetProArgHisSerLeuMetLysLys 530

4367 TGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATC... 4413
                ::::|||      |||:::::      :::   :::
 531 ..ThrIleAlaThrProCysArgThrLeuGluGluIleSerAspLeuAsp 546

4414 .........GTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGT 4454
                 |||:::   |||      :::   :::
 547 LeuAspAspPheAspAspLeuProAsnGluAlaSerAsn........... 559

4455 GAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAAC 4504
                                          |||      :
 560 ....................................GluProProS 563

4505 AGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGC 4554
     ::::::::   :::::::|||   :::::::::   |||   :::::::|||
 563 erAlaAsnGluGlnSerProAspAsnHisAlaGluThrThrThrArgGly 579

4555 AAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCA 4604
                       :::|||      |||:::||||||:::::
 580 .....................ValPhe...ProCysGluCysGlyThrGl 588

4605 GCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCAT 4654
     ::::   |||                    :::   |||::::::|||:
 588 uIleThrVal.....................AsnSerPheGlyArgAlaI 598

4655 TT..............ACCTTTGTCGATAACTTGAAAGGGAGGAGTGCG 4689
     ::                :::::::::|||:::::|||||||||   |||
 598 leGluValAlaGlyValAsnLeuThrAspHisMetLysGlyArgLeuAla 614

4690 GTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCA 4739
     :::|||:::|||:::   |||:::||||||:::|||:::|||   |||||
 615 AlaPheTyrSerArgAspGlyGlnGlyTyrSerTyrThrGlyTyrSerHi 631

4740 TGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATT. 4788
     |   |||:::||||||   ::::::|||:::   ::::::   |||
 631 sLysSerGlnGlyTrpLeuGluGlyLeuAspLysLeuIleGluAlaCysG 648

4789 ..AAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATG 4836
       :::   |||:::::::::::::::|||||||||||||||||:::::
 648 lyGluLysProThrThrTyrAsnGlnCysLeuValGlnLysTyrGluGln 664

4837 GGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGA 4886
     |||:::   :::   ||||||:::|||:::::      ||||||:::
 665 GlySerArgIleGlyPheHisSerAspGluGlnAlaIleTyrProLysGl 681

4887 TAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTA 4936
      |||   ||||||||||||||   :::|||:::::::::|||:::   |
 681 yAsnLysIleLeuThrValAsnAlaAlaGlySerGlyThrPheGlyIleL 698

4937 AGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGAC 4986
     |||||   ||||||           ::::::|||:::   :::||||||
 698 ysCysAlaLysGly......GluThrThrLeuAsnLeuGluAspGlyAsp 712
```

```
4987 TATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAA 5036
     ||||||   ||||||   |||||||||:::||||||   |||:::|||
 713 TyrPheGlnMetProSerGlyPheGlnGluThrHisLysHisAsnValVa 729

5037 CTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCG 5086
       ::::::         |||:::|||:::|||||||||:::|||
 729 lAlaValThrPro...ArgLeuSerPheThrPheArgSerThr..... 742

5087 TCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAG 5136
                                              |||::::::::::
 743 ..................................ValValAsnSerGln 747

5137 AAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAG 5186
     |||:::|||:::   |||::::::      |||:::::::::::   ::
 748 LysLysProAlaGluProGluLysLeuAsnGlnAsnAsnAlaCysProLy 764

5187 AACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTA 5236
     :         |||                        |||::::::||||||:
 764 sProSerAspPro.....................SerAsnAlaSerGlyL 774

5237 GGGAA..................GTCAAGGGGAGGTCGACATACTCGATA 5268
     :::::                  :::||||||:::::
 774 ysGlnHisLysLysThrHisProAlaLysGlyAsnGlu............. 786

5269 TGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAA 5318
                                          ::::::::::
 787 ....................................LysSerSerSe 790

5319 TCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGG 5368
     :|||         |||||||||      :::|||   |||   |||::::
 790 rPro......AsnLeuGluProLeuAspAlaPro...ThrValGluIleL 804

5369 TTAAA..............................ACCGGGACCTCT 5385
     ::|||                                  ||||||
 804 euLysLeuHisGlyPheThrAlaLeuThrProGlnHisAspGlyThrCys 820

5386 GAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAG 5435
     :::         |||   :::   |||          |||   :::::||
 821 GlnIleArgProValTyrPheAsnLys......AspIleHisLeuArgAr 835

5436 GACATACAGGGCGTTGCTTATG.........GCTAGA.........AATA 5467
     |                  |||         ||||||         ::::
 835 gLysAlaValLysThrAspMetSerProProAlaArgProPhePheAspL 852

5468 TTGCCGTCACTACCGCCGAAGGTGTTCTG......AAAGTACCTAATCAA 5511
     ::|||::::::      :::|||:::         |||:::   |||:::
 852 euAlaThrSerLeuHisArgGlyIleTyrThrHisLysIleAspAsnArg 868

5512 .......................GTTTATGAATCACTACCGGGCTTTCA 5537
                             |||   :::::|||   |||:::
 869 ArgAlaThrAlaTyrMetSerAspValLysAsnAsnLeuThrGlyLeuVa 885

5538 CGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCT 5587
      :::   |||         ||||||::      |||:::      :::|
 885 lLeuProLysLeuAspArgAspLeuLeu......SerSerTrpValAlaL 900

5588 TG.........CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAA 5628
     ||            :::|||::::::      :::   |||
 900 euAlaGluThrThrThrArgGluValAlaValLeuAlaIleHisGlyAla 916

5629 GGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGA 5678
     |||         |||::::::   ::::::   ::::::      :::
 917 GlyGlyAlaGlyLysSerArgAlaLeuGlnGluLeuLeuArgSerSerPr 933

5679 CAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATT 5728
      :::|||      :::|||:::|||   :::       :::|||||||
 933 oGluLeu......AlaAspSerIleAsnIleValValProThrIleAsnL 948

5729 TGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTT 5778
     ||   :::      |||:::           |||      :::
 948 euAlaAsnAspTrpLysAlaLysLeuProGlnMetAspProArgArgVal 964

5779 AAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGAC 5828
```

```
      :::|||:::      |||   :::                  |||::
 965 MetThrPheGlnLysAlaCysGluArgGlu...........CysLysSe 977

5829 GACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCA 5878
     :::|||:::   |||::::::   |||   |||      ::::::   :
 977 rValThrIlePheAspAspTyrGlyLysLeuPro......AlaGlyPheV 992

5879 TTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTG...... 5922
     ::   |||   ::::::   :::   ::::::::::::|||:::
 992 alAspAlaTyrLeuAlaIleLysValAsnValGluLeuAlaIleLeuThr 1008

5923 .........................AAGAAGAGAGATCCGAATTTGGA 5945
                               :::::::|||:::   ::::::::
1009 GlyAspGlnArgGlnSerThrHisHisGlnGluArgGluSerGlnIleSe 1025

5946 AGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCG 5995
     ::::|||:::|||      :::      |||:::   :::::::::   :
1025 rSerLeuGlnSerAsnIleAlaGlnPheSerLysTyrAlaAspTyrTyrL 1042

5996 TCAGGGGAATGTAT.................................. 6009
     ::::::::   :::
1042 euAsnAlaThrHisArgGlnProArgArgLeuAlaAsnProIleLysVal 1058

6010 ...............................AAAAGGGTTTTGGTGGA 6026
                                    |||      :::|||
1059 HisAlaGluArgGlnLeuGlyGlyAlaValLeuLysAlaAsnIleValPr 1075

6027 TGAGGTGCACATGATG.................CATCAAGGCTTACTA. 6057
      ::::::   |||:::                  |||:::||||||
1075 oAspLeuAlaMetValLeuValProAlaPheArgSerGlnSerLeuLeuT 1092

6058 ..CAACTAGGCGTCTTCGCAACC......GGCGCGTCGGAAGGCCTCTTT 6099
       :::||||||      |||         ::::::   :::||||||
1092 hrAspLeuGlyArgHisAlaMetThrTyrAlaGlyCysGlnGlyLeuThr 1108

6100 TTTGGAGACATAAATCAGATA..............CCATTCATAAACAG 6134
     ::::::   ::::::   |||               |||:::   :::
1109 LeuAsnHisLeuThrIleIleLeuAspLysAspThrProLeuCysSerAs 1125

6135 GGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAA 6184
      |||   ::::::            ::::::|||         :::||||
1125 pGluValLeuTyr............ThrAlaPheSerArgAlaSerGluS 1138

6185 GCGTTGTATACACT..............TCTAAATCGTACAGGTGTCCG 6219
     ||::::::::::::            :::   ::::::   :::
1138 erIleThrPheValAsnThrHisSerAspAsnProAlaPheLeuAlaLys 1154

6220 TTAGATGTTTGCTACTTGTTGTCCTCAATGACC......GTAAGGGGAAC 6263
     ||||||:::         |||:::::::::         ||||||
1155 LeuAspAlaThrProTyrLeuLysThrLeuIleSerTrpValArgGluAs 1171

6264 GGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAG 6313
      |||:::                  :::|||   |||   |||::::
1171 pGluGlu.....................AlaGlyAlaAspCysProAlaT 1181

6314 TAAGATCGCTGTCCAAAAGG...CCAATTGGAACCACTGATGACGTAGCT 6360
     :::::   |||   |||      |||      |||         ||||||
1181 hrGluProLeuValLysAspValProThrLysThrHisIleProValAla 1197

6361 GAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
     :::   ::::::::::         :::      :::|||   :::||
1198 AsnAspLysValGlnLeuGluGlyLysIleGluAlaMetGluAspLysAs 1214

6411 TATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAG 6460
     |   ::::::      :::|||:::   |||   :::   ::::::|||
1214 pThrArgGluLeuTrpSerGlyGluGluLysThrAsnLeuMetGlnThrG 1231

6461 TGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACG 6510
           :::                    |||   ||||||
1231 lnAspProVal.....................ValGlnLeuPheProHis 1240

6511 AAGAAAGCC...GATGACTCCCTATTCACTAAACAACCGCATATACTTGT 6557
     ::::::|||   |||::::::||||||
```

```
1241 GlnGlnAlaLysAspGluAlaLeuPhe..,..... 1249

6558 TGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAA 6607
                                                      !
1250 .....................................................L 1250

6608 AGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCA 6657
     ||:::         :::|||      |||      |||:::|||::::::
1250 ysIle......ThrIleGlyGluArgIleArgMetAlaThrProGluGln 1264

6658 GTATCCGACGCTTTGCTTCACACG 6681
        :::      |||   ||||||
1265 AsnAlaLysGlnLeuArgHisThr 1272
```

```
seq_name: SwissProt_39:RRPO_PMV

seq_documentation_block:
ID   RRPO_PMV       STANDARD;      PRT;  1547 AA.
AC   P20951;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA REPLICATION PROTEIN (176 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
DE   DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
OS   Papaya mosaic potexvirus (PMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89381685; PubMed=2778435;
RA   Sit T.L., Abouhaidar M.G., Holy S.;
RT   "Nucleotide sequence of papaya mosaic virus RNA.";
RL   J. Gen. Virol. 70:2325-2331(1989).
CC   -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC       POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13957; BAA03050.1; -.
DR   PIR; JQ0096; JQ0096.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW   Transferase.
FT   NP_BIND     822    829       ATP (POTENTIAL).
SQ   SEQUENCE   1547 AA;  176320 MW;  6CA4282C6A082622 CRC64;


alignment_scores:
            Quality:  404.00              Length:    798
              Ratio:  1.101                 Gaps:     35
 Percent Similarity:  45.990   Percent Identity: 23.559

alignment_block:
 US-09-301-906-3 x RRPO_PMV   ..

 Align seg 1/1  to: RRPO_PMV  from: 1  to: 1547

    4157 GTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGT 4206
         |||   |||   :::   :::|||      ::::::::::::      ||
     405 ValThrLeuThrTyrProThrLysThrPheAspAsnArgGlyTrpValVa 421

    4207 GAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACA 4256
         |:::   :::  ::::::|||   |||               :::||||
     421 lLysLeuGluAlaArgArgGlyTyrGluTrpPheAlaAspGluMetHisL 438

    4257 AACCGCGGTCAACGAAG.............................. 4273
         |||||::::::   :::
```


438 ysProLysGlyProGluLeuAsnLeuGluGluLysLysThrAspProAsp 454
4274 ...GCGACAGGGAGCCAGTACAG.....TCCAGTCTTGTGAGTTC 4310
455 AlaAlaSerTyrGluLysTyrLeuLysAlaLeuSerLeuLeuGlnLysGl 471
4311 GCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGA 4360
471 uProGluValMetGluAlaLysGluAlaGluAlaThrAspGluProGlnA 488
4361 AA...GAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACG 4407
488 rgProGluValLysGluGluGlnAlaGluAlaSerThrSerGlyArgAla 504
4408 CCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAA 4457
505 GluGluIleGlnGluAspProAlaThrLys..... 514
4458 GATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGG 4507
515 LysGlyLys..... 517
4508 TACAAGTCGAGCAGCCCAAA...CAGAGGAGTTTGACGATCAATGAAGGC 4554
518 GluGluProAsnProAsnArgAspLeu..... 526
4555 AAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCA 4604
527 LeuCysProCysGlyLeuHi 533
4605 GCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAAC.... 4650
533 sLeuLysIleLysAsn.....AlaGluPheProGluLeuP 545
4651 ..GCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTC 4698
545 roValLeuAspHisProAspHisLeuThrGlyArgLysAlaTrpPhePhe 561
4699 TCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATC 4748
562 SerLysAspGlyLysProTyrSerTyrThrGlyGlySerHisAlaSerAr 578
4749 AGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCA......ATTAAGT 4792
578 qGlyTrpProAsnTrpLeuGluLysIleLeuAlaAlaIleGluIleLysG 595
4793 ACCCA...AGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGT 4839
595 luProLeuProGluPheAsnGlnCysLeuValGlnGlnPheLysLeuGln 611
4840 GGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAA 4889
612 AlaAlaIleProPheHisArgAspAspGluProCysTyrProLysGlyHi 628
4890 CCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGT 4939
628 sGlnValLeuThrIleAsnHisSerGlyGluCysLeuThrGlnIleAlaC 645
4940 GCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTAT 4989
645 ysGlnLysGly......LysAlaSerIleThrMetGlyPheGlyAspTyr 659
4990 TTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTC 5039
660 TyrLeuSerProValGlyPheGlnGluSerHisLysHisAlaValSerAs 676
5040 CATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACT..... 5079
676 nThrThrGlyGlyArgValSerLeuThrPheArgCysThrValGlnGlnA 693
5079 5079
693 snLysPheAsnAspAspGlySerMetGluAlaLeuAspAsnLeuProTrp 709
5080 CGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
710 LysAlaTrpIleProLysLeuGlnAsnLeuGlyPheGlnGlyArgGlnLe 726
5109 GCAG.....TTAGCCGGCGGCGTGTCGGATGAGAAGT 5140
726 uGlnTyrAspProAsnGlyAlaLeuIleSerProIleGluGluIleArgS 743
5141 CACCAGGTGTTCCAAACCAGCAACCACAGAGC..... 5172
743 er.....MetProLysCysLysProGluGlyValProGluValValTyr 757
5173 CAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAA 5213
758 LysThrLeuAspGlyLeuAlaArgAlaProThrProTyrSerProAsnPr 774
5214 GGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACT 5263
774 oIleArgAlaArgAlaTyrThrSerAspValLysAsnCysArgIleGlyA 791
5264 CGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313
791 laLeuLeuArgGlnGlnGlyLys.....GluTrp..... 800
5314 GATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGA 5363
800 800
5364 AGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGT 5413
801 GlyCysArgPheAspAlaLeuValGluAlaGlyLys. 812
5414 ATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGA 5463
813 Arg 813
5464 AATATTGCCGTCACTACCGCCGAAGGTGTT...CTGAAAGTACCTAATCA 5510
814 GluLeuAlaIleSerValIleHisGlyAlaGlyGlySerGlyLysSerGl 830
5511 AGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATC 5560
830 nAlaLeuGlnThrLeu..... 835
5561 TCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTAC 5610
836 IleLysAspAsnPro... 840
5611 GTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGC 5660
841 GluLeuAspIl 844
5661 GGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGTTT 5710
844 eThrValValLeuProThrAsn..... 851
5711 TCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGC 5760
852 GluLeuArgLeuAspTrpLeuArgLysLeuProLysAlaPro 865
5761 ATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCC 5810
866 Gln.....GluLysPheLysThrPheGlu...LysAlaLeuLeuAlaPr 879
5811 CCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGTCAC 5860
879 oPro.....ThrProIleValIlePheAspAspTyrGlyLysLeuP 893
5861 CCAAT.....AGCACAGCCACCATT 5880
893 roAlaGlyTyrValGluAlaPheCysLeuTyrPheSerThrValGlnLeu 909

```
    5881 ACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAG 5930
                 :::|||:::                                 |||::
     910 IleIleLeuThrGlyAspCys.....................LysGl 918

    5931 AGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGG 5980
         ::::   :::   |||:::   ::::::|||||||||   :::
     918 nSerValHisHisGluSerAsnGluAsnAlaThrThrSerSer..... 932

    5981 TGGTAAACTTTATCGTCAGGGGAATGTATAAA......AGGGTTTTGGTG 6024
           ::::::   :::|||:::         :::      |||      :::
     933 ..IleGluProLeuValLysGluAlaSerGluLeuCysArgTyrTyrIle 948

    6025 GATGAGGTGCACATGATGCATCAAGGCTTACTA...CAACTAGGCGTCTT 6071
         :::   :::|||         :::   |||      :::||||||||::
     949 AsnAlaThrHisArgAsnLysLysAspLeuAlaAsnLysLeuGlyValTy 965

    6072 CGCA......ACCGGCGCGTCGGAA..................... 6090
         ::::      ||||||   :::|||
     965 rSerGluLysThrGlyLeuThrGluValThrHisGlyThrThrProIleP 982

    6091 ..GGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAG 6138
           ||||||   :::               :::|||   :::   ::::::
     982 roGlyLeuHisMet...........LeuValProSerLeuTyrLysLys 994

    6139 AAGGTGTTTAGG.......................ATGGATTGTGC 6161
         ::::::|||                          |||
     995 GlnAlaPheSerGluMetGlyHisLysValSerThrTyrAlaGlyCysGl 1011

    6162 TGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACA 6211
             ::::::      |||   :::::::::   |||:::::::::
    1011 nGlyIleThrAlaProLysIleGlnIleLeuLeuThrGluGluThrSerL 1028

    6212 GGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTA 6255
           |||      :::|||   |||         |||   :::|||
    1028 euCysSerArgGluValLeuTyrThrAlaLeuSerArgAlaVal 1042

seq_name: SwissProt_39:194K_TRVSY

seq_documentation_block:
ID   194K_TRVSY     STANDARD;      PRT;  1707 AA.
AC   P05080;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-JUL-1989 (Rel. 11, Last annotation update)
DE   POTENTIAL 194 KDA PROTEIN (PUTATIVE REPLICASE) [CONTAINS: 134 KDA
DE   PROTEIN].
OS   Tobacco rattle virus (strain SYM).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88034943; PubMed=3668507;
RA   Hamilton W.D.O., Boccara M., Robinson D.J., Baulcombe D.C.;
RT   "The complete nucleotide sequence of tobacco rattle virus RNA-1.";
RL   J. Gen. Virol. 68:2563-2575(1987).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A.
RA   Boccara M., Hamilton W.D.O., Baulcombe D.C.;
RT   "The organisation and interviral homologies of genes at the 3' end of
RT   tobacco rattle virus RNA1.";
RL   EMBO J. 5:223-229(1986).
CC   -!- SIMILARITY: THIS PROTEIN IS HOMOLOG TO PUTATIVE REPLICASE OF
CC       TOBACCO MOSAIC VIRUS (TMV) AND TRICORNAVIRUSES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06172; CAA29537.1; -.
DR   EMBL; D00155; BAA00110.1; -.
DR   PIR; S01865; S01865.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
FT   CHAIN         1   1187       134 KDA PROTEIN.
FT   CHAIN         1   1707       194 KDA PROTEIN.
SQ   SEQUENCE   1707 AA;  194328 MW;  F4DABDB6C7F51F04 CRC64;


alignment_scores:
           Quality:  338.50              Length:   1858
             Ratio:  0.486                 Gaps:     72
 Percent Similarity:  37.513   Percent Identity: 17.976

alignment_block:
 US-09-301-906-3 x 194K_TRVSY   ..

 Align seg 1/1  to: 194K_TRVSY  from: 1  to: 1707

    1333 GACGCAATGATTCGCGAAAAACCCTCA.................CATAG 1364
         |||   |||||||||||||:::   :::                   |||::
      50 AspAspMetIleArgGluArgLysThrLysAspProValIleValHisGl 66

    1365 GTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCT 1414
         :           |||:::   :::|||::::::||||||   |||::::
      66 u.........ValLeuSerGlnLysGluGlnAsnLysLeuMetGluIleT 80

    1415 TTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCA 1464
         ::||||||::::::|||   |||:::|||:::   :::      |||
      80 yrProGluPheAsnIleValPheLysAspAspLysAsnMetValHisGly 96

    1465 TTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTG 1514
         ||||||   |||   |||:::            ::::::      |||
      97 PheAlaAlaAlaGluArgLysLeuGlnAlaLeuLeuLeuLeuAspArgVa 113

    1515 TGGTAATGTGTGC...TTCTTCGATATTGGGGGGAGCTTCACGTATCATG 1561
              :::            ||||||||||:::::::::::   |
      113 lProAlaLeuGlnGluValAspAspIleGlyGlyGlnTrpSerPheTrpV 130

    1562 TCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTT 1611
         ||      |||:::   :::   |||   |||   |||   ||||||:::
      130 alThrArgGlyGluLysArgIleHisSerCysCysProAsnLeuAspIle 146

    1612 AAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGG 1661
         :::|||   :::|||:::|||::::::   :::||||||::::::   ||
      147 ArgAspAspGlnArgGluIleSerArgGlnIlePheLeuThrAlaIleGl 163

    1662 GGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAA............. 1698
         |   ::::::      ::::::      ::::::|||
      163 yAspGlnAlaArgSerGlyLysArgGlnMetSerGluAsnGluLeuTrpM 180

    1699 .....................GCGGCTTCAAAGTCTGTGTCTTACTGT 1725
                                    |||:::   ::::::      |||
      180 etTyrAspGlnPheArgLysAsnIleAlaAlaProAsnAlaValArgCys 196

    1726 AGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTT..... 1770
         ::::::      |||:::|||      |||         ||||||
      197 AsnAsnThrTyrGlnGlyCysThrCysArg.........GlyPheSerAs 210

    1771 ...............................ATGGTGGATGTGTACG 1786
                                        :::   ::::::||||
      210 pGlyLysLysLysGlyAlaGlnTyrAlaIleAlaLeuHisSerLeuTyrA 227

    1787 ATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTG 1836
         ||::::::   :::::::::      :::|||   :::|||   :::
      227 spPheLysLeuLysAspLeuMetAlaThrMetValGluLysLysThrLys 243

    1837 GTTTTCGACATAGCTCTTATGTTC...CCCGTGGAGTTGTTGTACGGTAA 1883
         |||         |||:::::::|||   |||   ::::::|||      ::
      244 ValValHisAlaAlaMetLeuPheAlaProGluSerMetLeuValAspGl 260
```

```
1884 CGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAGGTG 1933
     :|||                              |||   :::   :::|||
 260 uGly.....................ProLeuProSerValAspGlyT 269

1934 ATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAACATTCC 1983
       |||:::      |||           |||::::::|||
 269 yrTyrMetLysLysAsn...........GlyLysIleTyr......... 278

1984 TTCTCTAACGTAAGCGGGTTTTTC.........ACCTTTTCTTATGTACG 2024
                   ||||||           :::||||||||||:::::
 279 ............PheGlyPheGluLysAspProSerPheSerTyrIleHi 291

2025 CACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTG 2074
     :      :::                    ||||||:::   |||
 291 sAspTrpGlu.................GluTyrLysLysTyrLeuLeu. 301

2075 GTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACTGAG 2124
                                                  |||
 302 .....................................GlyLysPro 304

2125 GTTACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTT 2174
     |||:::||:::                    |||:::   :::   ||
 305 ValSerTyrGln.....................GlyAsnValPheTyrPh 314

2175 CATACCT...GTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCC 2221
     |   |||      |||   |||:::::::::   |||:::
 314 eGluProTrpGlnValArgGlyAspThrMetLeuPheSer..... 327

2222 TCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATA 2271
                          ||||||     |||
 328 ......................IleTyrArgIleAla............ 332

2272 GGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTC 2321
                                        |||:::   |||
 333 ....................................GlyValProArg.. 336

2322 GCAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTG 2371

 336 ................................................ 336

2372 ATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAG 2421

 336 ................................................ 336

2422 GCGATTAAGGATAGGGCGAAGAGTATTCGCTCCTATAACTTTATAAAAGC 2471
                     :::|||:::   |||   ::::::   :::
 337 ................ArgSerLeuSerSerGlnGluTyrTyrArg.. 346

2472 CAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAG 2521

 346 ................................................ 346

2522 GCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCAC 2571

 346 ................................................ 346

2572 GATAACTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCAT 2621
                                        |||   :::||
 347 ....................................ArgIleTyrIl 350

2622 CCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTT 2671
     |   :::      :::   |||||||||   |||
 350 eSerArgTrpGluAsnMetValValValProIle............... 361

2672 CAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTC 2721
                              ||||||:::   ::::::|||
 362 ........................PheAspLeuValGluSerThr... 368

2722 GGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGT 2771
                                        |||   :::||
 369 ....................................ArgGluLeuVa 372
```

```
2772 GGAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAAC 2821
     |:::   ::::::   |||   |||   :::
 372 lLysLysAspLeuPheValGluLysGlnPheMet.............. 383

2822 CCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGG..... 2865
                        |||:::         :::|||
 384 ................AspLysCysLeuAspTyrIleAlaArgLeuSer 394

2866 ............AACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCT 2903
                  ::::::|||||||||
 395 AspGlnGlnLeuThrIleSerAsnVal..................... 403

2904 TGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAG 2953
                       :::   :::||||||   :::      |||
 404 ............LysSerTyrLeuSerSerAsnAsnTrpValLeuPheI 416

2954 CTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCC 3003
       :::||||||||||:::|||
 416 leAsnGlyAlaAlaValLys........................... 422

3004 ACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGA 3053
                       ::::::      ::::::::::::   |||::
 423 ..................AsnLysGlnSerValAspSerArgAspLeuGl 433

3054 TGCTCTGGTGCGCACA.........GAGAATGCTATAAACTCACCGGTGG 3094
     :   |||:::::::|||         ::::::   :::      ||||||
 433 nLeuLeuAlaGlnThrLeuLeuValLysGluGlnValAlaArgProVal. 449

3095 GCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACG 3144
                       |||         :::   :::::::::   |||
 450 ..............MetArgGluLeuArgGluAlaIleLeuThrGluThr 461

3145 AAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATT 3194
     |||   :::          :::||||||:::|||
 462 LysProIle.........ThrSerLeuThrAsp.................. 469

3195 GCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCC 3244
                    :::::::|||::::::      :::|||
 470 .............ValLeuGlyLeuIleSerArgLysLeu.......... 478

3245 CAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTAC 3294
             ||||||                              ::::::
 479 ........TrpLys.............................GlnPhe 482

3295 GCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGG 3344
     |||                                 ::::::   |||:::||
 483 Ala............................AsnLysIleAlaValGl 489

3345 TGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCG 3394
     ||||:::   |||::::::                    |||:::|||:::
 489 yGlyPheValGlyMetVal.................GlyThrLeuIle. 499

3395 TGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCT 3444
             ||||||              ::::::|||
 500 ........GlyPheTyrProLysLysValLeuThr.............. 508

3445 AAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCAT 3494

 508 ................................................ 508

3495 TTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACT 3544
                                  ||||||         |||:::
 509 ........................TrpAlaLysAspThrProAsnGly. 516

3545 TAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCG 3594

 516 ................................................ 516

3595 AATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGA 3644
                                                  |||||
 517 ..............................................ProGl 518

3645 GGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGC 3694
```

```
     |:::   ::::::::::::      ::::::::::::::::   :::::::
 518 uLeuCysTyrGluAsnSerHisLysThrLysValIleValPheLeuSerV 535

3695 TA..............GGGGTATCTTTGATCATAAGGGGCTTGCTTAAC 3729
     ::              |||:::::::|||:::   |||   :::   :::
 535 alValTyrAlaIleGlyGlyIleThrLeuMetArgArgAspIleArgAsp 551

3730 GACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAG...GGGAGGAA 3776
       :::   :::|||:::   :::      ::::::   |||   ::
 552 GlyLeuValLysLysLeuCysAspMetPheAspIleLysArgGlyAlaHi 568

3777 TGTGTACGAT......GAGACGCTTAGGTATTACCGGGACTTTGACTATG 3820
     :|||  |||      :::      |||||||||
 568 sValLeuAspValGluAsnProCysArgTyrTyr..............  579

3821 ACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGAC 3870
                                                    :::
 580 ..............................................Glu 580

3871 GATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAA 3920
       ||||||      :::|||      |||:::
 581 IleAsnAspPhePheSerSerLeuTyrSerAla..............  591

3921 TGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAG 3970
                       |||   :::||||||:::
 592 ..................SerGluSerGlyGluThr.............  597

3971 AGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTT 4020
                                                    |||
 598 ..............................................Val 598

4021 GCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGG 4070
       |||:::   :::   ::::::   :::::::|||
 599 LeuProAspLeuSerGluValLysAlaLysSerAsp..............  610

4071 TACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACG 4120
                       ::::::::::::   |||:::   :::|||:
 611 ................LysLeuLeuGlnGlnLysLysGluIleAlaAspG 622

4121 ATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTG 4170
     ::   :::      :::   |||:::   ::::::   :::|||::::::
 622 luPheLeuSerAlaLysPheSerAsnTyrSerGlySerSerValArgThr 638

4171 GACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACA 4220
     :::              |||   ||||||||||||::::::   :::
 639 Ser..............ProProSerValValGlySerSerArgSerGl 650

4221 AGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACG 4270
                       |||:::|||:::
 650 yLeuGlyLeuLeuLeuGluAspSerAsnVal...................  660

4271 AAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCT 4320

 660 ..............................................  660

4321 GATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTG.. 4368
               :::||||||:::::::|||   :::::::::::::|||
 661 ............LeuThrGlnAlaArgValGlyValSerArgLysValAs 673

4369 .AAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCG 4417
         :::|||:::            :::|||||||
 673 pAspGluGluIleMetGluGlnPheLeuSerGly..............  684

4418 ATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGAC 4467
                                                :::||||||
 685 ..........................................LeuIleAsp 687

4468 AAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGA 4517
               :::   :::|||:::::::      :::::::::   ||
 688 ThrGluAlaGluIleAspGluValValSerAlaPheSerAlaGluCysGl 704

4518 GCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAAC 4567
     |:::                        :::|||   ::::::   |||
```

```
 704 uArg.......................GlyGluThrSerGlyThrLysV 713

4568 AGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTAC 4617
       ||||||         |||
 713 alLeuCysLysProLeuThr............................. 719

4618 AACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGA 4667

 719 ..............................................  719

4668 TAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGT 4717

 719 ..............................................  719

4718 ATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTA 4767
                                                |||   ::::::
 720 ...................................ProProGlyPhe 723

4768 GAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTT 4817
     |||::::::|||   |||:::|||                       ||
 724 GluAsnValLeuProAlaValLys.....................ProLe 733

4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACG 4867
     ||||:::|||   |||
 733 uValSerLysGlyLys................................ 738

4868 AGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGG 4917
                                       ||||||:::   |||
 739 ...............................ThrValLysArgVal... 743

4918 AAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCAT 4967

 743 ..............................................  743

4968 AAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTT....... 5010
                     |||||||||   :::      |||
 744 ................AspTyrPheGlnValMetGlyGlyGluArgLeuP 755

5011 ..CAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATC 5058
       :::|||      :::   :::   ||||||:::|||      |||
 755 roLysArgProValValSerGlyAspAsnSerValAspAlaArgArgGlu 771

5059 AGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
       |||   :::      ::::::||||||                :::::::
 772 PheLeuTyrTyrLeuAspAlaGluArgValAlaGlnAsnAspGluIleMe 788

5109 GCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACC 5158
     :::|||
 788 tSerLeu.....................................T 791

5159 AGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGG 5208
       :::      |||:::|||:::   ||||||                |||
 791 yrArgAspTyrSerArgGlyValIleArgThr..............Gly 802

5209 GGCAAGGCTCTATCTGAGGGAAGTGGT.........AGGGAAGTCAAGGG 5249
     |||:::         :::|||   |||           |||:::|||::
 803 GlyGlnAsnTyrProHisGlyLeuGlyValTrpAspValGluMetLysAs 819

5250 GAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTG 5299
     :                  ||||||   :::      |||   :::
 819 n.................TrpCysIleArgProValValThrGluHisA 830

5300 AGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACC 5349
       ::::::                        ::::::||||||
 830 laTyrVal.......................PheGlnProAsp...... 836

5350 CCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGT 5399
                                                    :::
 837 ..................................LysArgMetAs 840

5400 GGAGTACTTGAAGTATCTGGCTATAGGCATT...GAGAGGACATACAGGG 5446
       ::::::      ||||||   :::::::::   ||||||
 840 pAspTrpSerGlyTyrLeuGluValAlaValTrpGluArg.......... 853
```

```
 5447 CGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTG 5496

  853 ..................................................  853

 5497 AAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAA 5546

  853 ..................................................  853

 5547 GTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGA 5596
                                          |||:::   |||:
  854 .................................GlyMetLeuValA 858

 5597 GAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGC 5646
      ::|||:::
  858 snAspPhe..................................... 860

 5647 AAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGA 5696
               ||||||      ::::::|||   ::::::|||||||
  861 ............AlaValGluArgMetSerAspTyrValIleValCysAs 873

 5697 CGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAG 5746
      |:::      :::   ::::::   :::   ::::::      |||::::
  873 pGlnThrTyrLeuCysAsnAsnArgLeu...IleLeuAspAsnLeuSerA 889

 5747 TCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAG 5796
      ::         |||   |||:::   ||||||                ::::
  889 laLeuAspLeuGlyProValAsnCysSerPhe..............Glu 900

 5797 TGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGA 5846
            ::::::   ||||||   |||||||:::|||   :::|||:::
  901 LeuValAspGlyValProGlyCysGlyLysSerThrMetIleValAsn.. 916

 5847 ATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA. 5895
               |||   |||      :::   ::::::   :::::::|||
  917 ..........SerAlaAsnProCysValAspValValLeuSerThrGlyA 930

 5896 ...............................AGTTCTGAGGACATAAAT 5913
                                     :::|||:::   :::
  930 rgAlaAlaThrAspAspLeuIleGluArgPheAlaSerLysGlyPhePro 946

 5914 ATGGCGGTGAAGAACAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTAC 5963
            :::|||:::|||                          :::
  947 CysLysLeuLysArgArg..........................ValLy 954

 5964 CACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAA 6013
       ||||||:::|||   :::::::::   :::      ::::::
  954 sThrValAspSerPheLeuMetHisCysValAspGlySerLeuThrGlyA 971

 6014 GGGTTTTG...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAA 6060
        ||||||      ||||||:::   |||   |||   |||::::::
  971 spValLeuHisPheAspGluAlaLeuMetAlaHisAlaGlyMetValTyr 987

 6061 CTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACAT 6110
      :::   :::      :::||||||:::::   :::      ||||||
  988 PheCysAlaGlnIleAlaGlyAlaLysArgCysIleCysGlnGlyAspGl 1004

 6111 AAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTG 6160
       |||||||||   |||      |||   :::          :::      :
 1004 nAsnGlnIleSerPheLysProArgValSerGlnValAspLeuArgPheS 1021

 6161 CTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTAC 6210
      ::   :::|||   |||   :::   |||:::          ::::::|||
 1021 erSerLeuValGlyLysPheAspIleValThrGluLysArgGluThrTyr 1037

 6211 AGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGG 6260
      |||   |||   ||||||:::   :::|||:::
 1038 ArgSerProAlaAspValAlaAlaValLeuAsn..................  1048

 6261 AACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAG 6310
             |||   |||         |||   :::         :::   :
 1049 .......LysTyrTyrThrGlyAspValArgThr......HisAsnAlaT 1061

 6311 TAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCT 6360
      :::::::|||::::::   |||   |||   :::   :::::::|||:::
 1061 hrAlaAsnSerMetThrValArgLysIleValSerLysGluGlnValSer 1077

 6361 GAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
                :::      |||:::   :::   |||   ||||||:::::
 1078 LeuLysProGlyAlaGlnTyrIleThrPheLeuGlnSerGluLysLysG  1094

 6411 TATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAG 6460
      ::::   :::   |||      :::         :::   |||   ||||
 1094 uLeuValAsnLeuLeuAlaLeuArgLysValAlaAlaLysValSerThrV 1111

 6461 TGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACG 6510
      ||||||||:::||||||:::|||||||:::||||||||||||   ||||||
 1111 alHisGluSerGlnGlyGluThrPheLysAspValValLeuValArgThr 1127

 6511 AAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGG 6560
      |||   :::||||||||||:::         :::   ::::::::::|||::
 1128 LysProThrAspAspSerIleAlaArgGlyArgGluTyrLeuIleValAl 1144

 6561 TTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGT 6610
      :|||||||||||||||:::||||||||||||   ::::::::
 1144 aLeuSerArgHisThrGlnSerLeuValTyrGluThrValLys....... 1158

 6611 TGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCAGTA 6660
        :::|||   |||:::      |||   :::::::::      ::::::
 1159 ..GluAspAspValSerLysGluIleArgGluSerAla......AlaLeu 1172

 6661 TCCGACGCTTTGCTTCACACGTTC 6684
      :::   |||   |||      |||
 1173 ThrLysAlaAlaLeuAlaArgPhe 1180
```

```
seq_name: SwissProt_39:RRPO_SHMV

seq_documentation_block:
ID   RRPO_SHMV      STANDARD;      PRT;  1629 AA.
AC   P89202; Q88624;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (128 KDA PROTEIN)].
OS   Sunn-hemp mosaic virus (SHMV) (TMV strain cowpea).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE OF 1-1536 FROM N.A.
RX   MEDLINE=97093421; PubMed=8938983;
RA   Silver S., Quan S., Deom C.M.;
RT   "Completion of the nucleotide sequence of sunn-hemp mosaic virus: a
RT   tobamovirus pathogenic to legumes.";
RL   Virus Genes 13:83-85(1996).
RN   [2]
RP   SEQUENCE OF 1538-1629 FROM N.A.
RX   MEDLINE=83064542; PubMed=6292867;
RA   Meshi T., Ohno T., Okada Y.;
RT   "Nucleotide sequence of the 30K protein cistron of cowpea strain of
RT   tobacco mosaic virus.";
RL   Nucleic Acids Res. 10:6111-6117(1982).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR LYS-1130 AND GLN-1132.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U47034; AAB38492.1; -.
DR   EMBL; J02413; AAA46585.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1629       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1130       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     849    856       ATP (POTENTIAL).
SQ   SEQUENCE   1629 AA;  186513 MW;  C928BC9AF95FE8A5 CRC64;


alignment_scores:
           Quality:  337.00              Length:   1898
             Ratio:  0.504                Gaps:     73
 Percent Similarity:  35.248   Percent Identity: 18.335

alignment_block:
 US-09-301-906-3 x RRPO_SHMV   ..

 Align seg 1/1  to: RRPO_SHMV  from: 1  to: 1629

    1279 TCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCTCA 1328
         |||:::|||:::::      :::|||:::   |||        :::
       4 SerThrLeuIleAsnLysAlaGlnThrAsnSerCysGlyAspValGlyVa 20

    1329 AAGGGACGCAATGATTCGCGAA..........................  1350
             |||   :::   |||:::
      20 lValAspLeuLeuLysArgLysValTyrAspAspThrValLysThrMetG 37

    1351 ..............AAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCG 1386
                     :::   ::::::|||   :::      |||   |||
      37 lnGlyLeuAspArgArgAlaLysTyrArgLeuAsnGlnCysLeuGlyPro 53

    1387 CGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTT 1436
         ::::::   :::   :::|||      :::||||||::::::|||:::||
      54 GluGlnCysArgThrValArgGlyGlyTyrProGluPheGlnIleGluPh 70

    1437 CTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCT 1486
         |:::   :::   ::::::||||||   :::|||   ::::::|||:::
      70 eThrGlyAlaSerAsnThrSerHisAlaMetAlaAlaGlyLeuArgGlyL 87

    1487 GT...TTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTC 1533
              ::::::   :::::::::            ::::::      :::
      87 euGluLeuGluTyrLeuTyrThrLeuValProTyrGlyAlaValSerTyr 103

    1534 GATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTG 1583
         ||||||||||||:::|||      |||:::      |||:::
     104 AspIleGlyGlyAsnPheProAlaHisMetMetLysGlyArgSerTyrVa 120

    1584 TCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATCA 1633
          |||   |||||||||:::||||||:::::|||:::   |||:::   |
     120 lHisCysCysAsnProAlaLeuAspAlaArgAspLeuAlaArgAsnGluA 137

    1634 ATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGT 1683
         ||      :::   :::                              :::
     137 snTyrArgIleSerIle..............................Glu 143

    1684 GACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTGTAGT..... 1728
         :::   |||::::::      :::||||||   :::||||||:::
     144 AsnTyrLeuSerArgPheGluAspLysSerGlyAspTyrCysGlnTrpGl 160

    1729 ................................CGAGAATCGCAGAACT 1744
                                        |||   ::::::   |
     160 nArgLysLysProLysValSerLysProLeuProArgTyrGlnLysAlaC 177

    1745 GCGATTCTAGAGCCGAT.............................. 1761
         ||   :::|||   :::
     177 ysPheAspArgTyrAsnGluAspProGluHisValThrCysSerGluThr 193

    1762 ..............................................GC 1763
                                                      ||
      194 PheGluLysCysArgIleSerProProAlaGluArgAspAspIleTyrAl 210

    1764 GGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGG 1813
         |      :::   ::::::|||||||||      |||:::::::::   |
      210 aThrSerLeuHisSerLeuTyrAspIleProTyrGlnAsnLeuGlyProA 227

    1814 CTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCC 1863
         ||:::   :::|||         |||:::      |||:::   |||
      227 laLeuAlaArgLysArgIleLysValLeuHisAlaAlaPheHisPheSer 243

    1864 GTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATAC 1913
            :::||||||   |||   :::|||      |||   ::::::
      244 GluAspLeuLeuLeuGlyAlaSerGluGlyLeuLeuThrGlnIleGlyGl 260

    1914 GTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAAT...GTTGGTCAGT 1960
               :::|||:::||||||   |||::::::::::   :::   :::
      260 yThrPheGlnArgAsnGlyAspValLeuThrPheSerPheLeuAspGluS 277

    1961 GTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACC 2010
            :::   :::|||   ||||||||   ||||||      :::   |||
      277 erSerLeuIleTyrThrHisSerPheArgAsnValPheGluTyrValThr 293

    2011 TTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGA 2060
         ::::::      :::   :::      ::::::      |||::::::
      294 ArgThrPhePheValAlaCysAsnArgTyrAlaTyrMetLysGluPheAr 310

    2061 GGGATACCGT..............TGTGGTTACCATCATCTC...ACTA 2092
         ::::   |||               |||::::::   ::::::   |||
      310 gSerArgArgValAspThrValPheCysSerPheIleArgIleAspThrT 327

    2093 TGTGTAGGGCTCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTG 2142
           |||                                 |||||||||:::
      327 yrCys............................LeuTyrArgSerVal 333

    2143 GTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGG 2192

      333 ..............................................  333

    2193 TTCTAGTGTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACA 2242
                   ||||||            |||   |||
      334 .............PheLys............AspCysAsp.......... 338

    2243 GGATCTATTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACG 2292

      338 ..............................................  338

    2293 TTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTAC 2342
           |||::::::   :::|||:::
      339 ...GluHisValPheAlaAlaMet.......................... 345

    2343 AGGGAGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGT 2392
                                                  ||||||   |
      346 .....................................AspAspAlaT 349

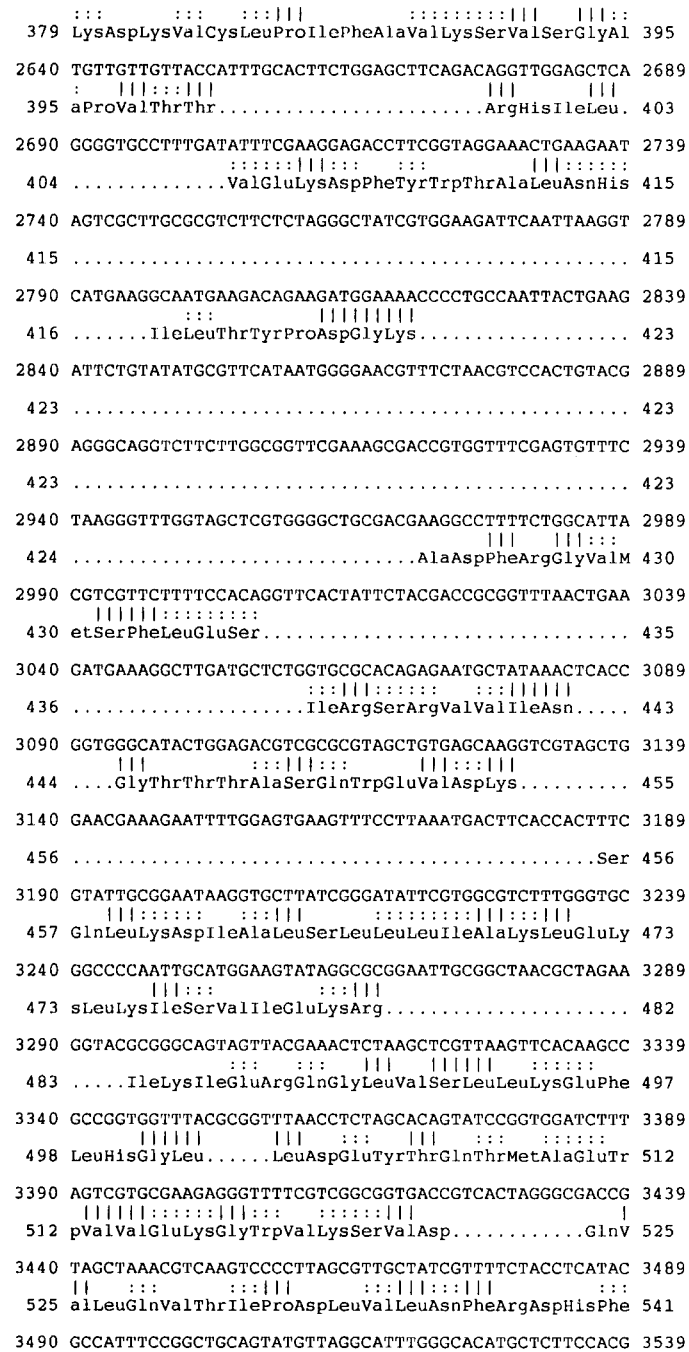
    2393 GGGGTTTA...GTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCG 2439
         ||   :::            ||||||   :::|||   :::
      349 rpGluPheLysLysLysArgValMetLeuGluAlaSerArgProIlePhe 365

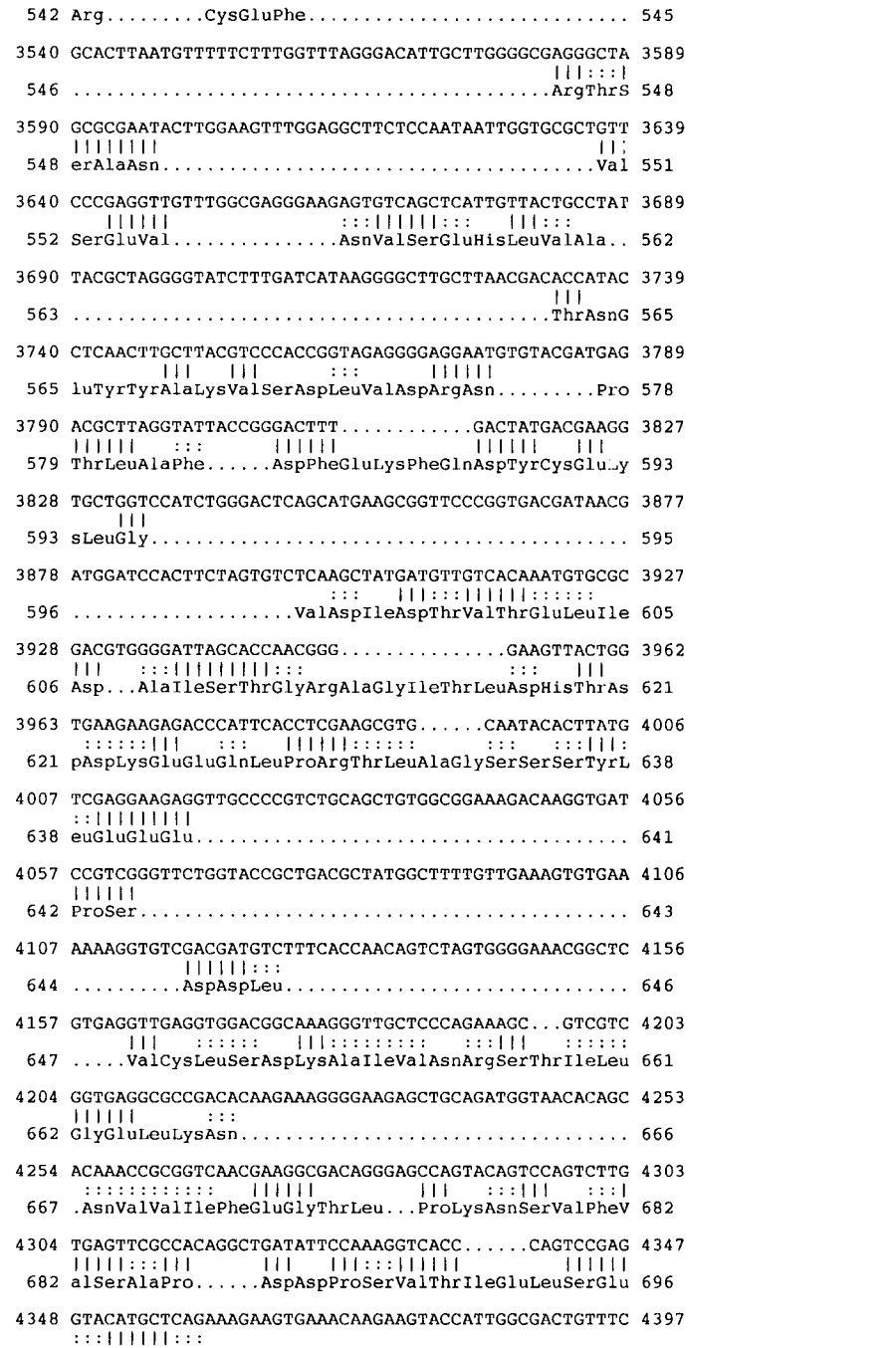
    2440 AAGAGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGC 2489
         :::::::::   :::::::|||
      366 AsnAspValAlaGlnPheAsnValTyr....................... 374

    2490 CGGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACG 2539
                                                  |||   ||||
      375 ....................................PheProAsnA 378

    2540 CAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTG 2589
         ||
      378 la.............................................. 378

    2590 GAGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTC 2639
```


379 LysAspLysValCysLeuProIlePheAlaValLysSerValSerGlyAl 395
2640 TGTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCA 2689
395 aProValThrThr.....ArgHisIleLeu. 403
2690 GGGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAAT 2739
404 ValGluLysAspPheTyrTrpThrAlaLeuAsnHis 415
2740 AGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGT 2789
415 415
2790 CATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAG 2839
416 IleLeuThrTyrProAspGlyLys..... 423
2840 ATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACG 2889
423 423
2890 AGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTC 2939
423 423
2940 TAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTA 2989
424 AlaAspPheArgGlyValM 430
2990 CGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAA 3039
430 etSerPheLeuGluSer..... 435
3040 GATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACC 3089
436 IleArgSerArgValValIleAsn..... 443
3090 GGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTG 3139
444 GlyThrThrThrAlaSerGlnTrpGluValAspLys..... 455
3140 GAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTTTC 3189
456 Ser 456
3190 GTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGC 3239
457 GlnLeuLysAspIleAlaLeuSerLeuLeuLeuIleAlaLysLeuGluLy 473
3240 GGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAA 3289
473 sLeuLysIleSerValIleGluLysArg..... 482
3290 GGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCC 3339
483 IleLysIleGluArgGlnGlyLeuValSerLeuLeuLysGluPhe 497
3340 GCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTT 3389
498 LeuHisGlyLeu.....LeuAspGluTyrThrGlnThrMetAlaGluTr 512
3390 AGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCG 3439
512 pValValGluLysGlyTrpValLysSerValAsp.....GlnV 525
3440 TAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATAC 3489
525 alLeuGlnValThrIleProAspLeuValLeuAsnPheArgAspHisPhe 541
3490 GCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCACG 3539


542 Arg.....CysGluPhe..... 545
3540 GCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTA 3589
546 ArgThrS 548
3590 GCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTT 3639
548 erAlaAsn.....Val 551
3640 CCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTAT 3689
552 SerGluVal.....AsnValSerGluHisLeuValAla.. 562
3690 TACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATAC 3739
563 ThrAsnG 565
3740 CTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAG 3789
565 luTyrTyrAlaLysValSerAspLeuValAspArgAsn.....Pro 578
3790 ACGCTTAGGTATTACCGGGACTTT.....GACTATGACGAAGG 3827
579 ThrLeuAlaPhe.....AspPheGluLysPheGlnAspTyrCysGlu..y 593
3828 TGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACG 3877
593 sLeuGly..... 595
3878 ATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGC 3927
596 ValAspIleAspThrValThrGluLeuIle 605
3928 GACGTGGGGATTAGCACCAACGGG.....GAAGTTACTGG 3962
606 Asp...AlaIleSerThrGlyArgAlaGlyIleThrLeuAspHisThrAs 621
3963 TGAAGAAGAGACCCATTCACCTCGAAGCGTG.....CAATACACTTATG 4006
621 pAspLysGluGluGlnLeuProArgThrLeuAlaGlySerSerSerTyrL 638
4007 TCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGAT 4056
638 euGluGluGlu..... 641
4057 CCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAA 4106
642 ProSer..... 643
4107 AAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTC 4156
644 AspAspLeu..... 646
4157 GTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGC...GTCGTC 4203
647 ValCysLeuSerAspLysAlaIleValAsnArgSerThrIleLeu 661
4204 GGTGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGC 4253
662 GlyGluLeuLysAsn..... 666
4254 ACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTG 4303
667 .AsnValValIlePheGluGlyThrLeu...ProLysAsnSerValPheV 682
4304 TGAGTTCGCCACAGGCTGATATTCCAAAGGTCACC.....CAGTCCGAG 4347
682 alSerAlaPro.....AspAspProSerValThrIleGluLeuSerGlu 696
4348 GTACATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTC 4397
697 LeuHisAlaArg..... 700

```
4398 GGGCGCCACGCCAATCGTCGAT.....GAGAAACCCGCCCCAA 4435
              |||:::   |||          :::||||||:::
 701 .........ProValSerAspPheLeuSerMetGlnLysProValAsnI 714

4436 GTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCT 4485
       |||   |||      |||:::|||
 714 leValTyrThrGlyGluValGlnIle..................... 722

4486 CATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAG 4535

 722 ..................................................  722

4536 TTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAA 4585

 722 ..................................................  722

4586 CGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCC 4635
       |||:::            ||||||   |||   :::|||:::::::::
 723 ..CysGlnMetGlnAsnTyrLeuAsp...TyrLeuSerAlaSerLeuVal 737

4636 ACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAG 4685
     :::   :::||||||         ::::::::::::         :::
 738 AlaCysIleSerAsnLeuLysLysTyrLeuGlnAspGlnTrpLeuAsnPr 754

4686 TGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTA 4735
     :::      |||:::|||:::|||                      :
 754 oGlyGluLysPheGlnLysIleGlyVal.................TrpA 765

4736 GCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCA 4785
     :::::::::::::   |||
 765 spAsnLeuAsnAsnLysTrp........................... 771

4786 ATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGAT 4835
     |||      |||                          ||||||   |||
 772 IleValValPro.....................GlnLysLysLysTy 780

4836 GGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAG 4885
       :::   |||:::        :::|||   ::::::
 780 rAlaTrpGlyLeuAlaAlaAspValAspGlyAsnGln............ 792

4886 ATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACT 4935

 792 ..................................................  792

4936 AAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGA 4985
                     |||:::::::::::::|||
 793 .................LysThrValIleLeuAsnTyr........... 799

4986 CTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAA 5035

 799 ..................................................  799

5036 ACTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC 5085

 799 ..................................................  799

5086 GTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGA 5135
                                                  |||||
 800 .............................................AspGl 801

5136 GAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCA 5185
     |         |||:::|||
 801 uHis......GlyMetPro.............................. 805

5186 GAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGT 5235

 805 ..................................................  805

5236 AGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTA 5285
                                                 |||:::::::||
 806 ..............................IleLeuGluLysSerTy 811
```

```
5286 CGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTG 5335
     |||||||
 811 rValArgLeu..................................... 814

5336 AACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCT 5385
                                      ||||||:::|||   |||
 815 ............................ValValSerThrAspThrTyr 821

5386 GAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAG 5435
       :::|||||||:::   |||   ||||||
 822 LeuPheThrValValSerMetLeuGlyTyrLeuArg.............. 833

5436 GACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCG 5485

 833 ..................................................  833

5486 AAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTT 5535

 833 ..................................................  833

5536 CACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGG 5585
                                        |||      |||
 834 ................................HisLeuAspGln..... 837

5586 CTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCT 5635

 837 ..................................................  837

5636 TTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTG 5685
               |||          |||:::::::::|||   |||::::::
 838 ...........LysLysProThrAlaThrIleThrLeuValAspGlyVal 850

5686 TTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGG 5735

 850 ..................................................  850

5736 TGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCT 5785

 850 ..................................................  850

5786 TCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACG 5835
                                  ||||||   |||||||||
 851 ........................ProGlyCysGlyLysThrGlnGlu 858

5836 TTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGC 5885
     :::::::::::|||       :::   ||||||          :::
 859 IleLeuSerArgPhe......AspAlaAsnSerAspLeuIleLeu..... 871

5886 TAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATC 5935
       |||   :::   |||       :::|||       ::::::   :::
 872 ....ValGlnGlyArgGluAlaCysGluMetIleArgArgArgAlaAsnA 887

5936 CGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTA 5985
       |||:::   |||      |||||||||      |||   |||:::
 887 spAsnValProGly......SerAlaThrLysGluAsnValArgThrPhe 901

5986 AACTTTATCGTCAGG.............GGAATGTATAAAAGGGTTTTGGT 6023
     :::   :::|||               |||   :::|||   :::   ||
 902 AspSerPheValMetAsnArgLysProGlyLysPheLysThrLeuTrpVa 918

6024 GGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCG 6073
     |||||||      |||:::|||   ||||||:::::::::   :::
 918 lAspGluGlyLeuMetValHisProGlyLeuIleAsnPheCysIleAsnI 935

6074 CAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCA 6123
       :::   :::|||:::      :::|||||||||   :::|||||||||
 935 leSerCysValSerSerValTyrIlePheGlyAspArgLysGlnIlePro 951

6124 TTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCC 6173
     ||||||||||||         |||   :::         :::
 952 PheIleAsnArgValMetAsnPheSerIleProAspAsnLeuAlaLysLe 968

6174 AAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAG 6223
```

```
            :::::::::|||      :::   :::   ||||||||||||
     968 uTyrTyrAspGluIleValSerArgAspThrThrLysArgCysProLeuA 985

    6224 ATGTTTGCTACTTGTTGTCCTCA.................ATGACCGTA 6255
         |||||   ::::::|||:::|||                 |||:::
     985 spValThrHisPheLeuAsnSerValTyrGluLysArgValMetSerTyr 1001

    6256 AGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAA 6305
           ::::::::::::            |||::::::|||||||||   ||
    1002 SerAsnValGlnArgSerLeuGluCysLysMetIleSerGlyLysAlaLy 1018

    6306 ACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACG 6355
         |                                      :::|||
    1018 s.....IleAsnAspT 1022

    6356 TAGCTGAAATAAACGCTGAC...GTGTACTTGTGCATGACCCAGTTGGAG 6402
           :::|||   |||:::         |||   :::||||||   :::
    1022 yrArgSerIleLeuAlaGluGlyLysLeuLeuThrPheThrGlnGluAsp 1038

    6403 AAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGT 6452
         |||:::   :::         ||||||:::     |||:::      ||
    1039 LysGluTyrLeu.........LeuLysAlaGlyPheLysAsp......Va 1050

    6453 GATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGT 6502
         |   |||||||||||||||||||||:::|||:::   ||||||   |||:
    1050 lAsnThrValHisGluAlaGlnGlyGluThrTyrArgAspValAsnLeuI 1067

    6503 TTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAA...CAACCGCAT 6549
         ::|||:::      :::      ::::::   :::      :::||||||
    1067 leArgValThrAlaThrProLeuThrIleValSerAlaGlySerProHis 1083

    6550 ATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCT 6599
         ::: ' |||:::|||||||||||||||||::   :::||||||   :::::
    1084 ValThrValAlaLeuSerArgHisThrAsnArgPheValTyrTyrThrVa 1100

    6600 GAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCAC 6649
         :          :::   |||   |||   |||   ::::::      :::
    1100 l.........ValProAspValValMetThrThrValGln......LysT 1112

    6650 CTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGGCT 6693
           |||   ||||||:::   ||||||      :::|||   |||
    1112 hrGlnCysValSerAsnPheLeuLeuAspMetTyrAlaValAla 1126

seq_name: SwissProt_39:V125_AMVLE

seq_documentation_block:
ID   V125_AMVLE     STANDARD;      PRT;  1126 AA.
AC   P03589;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   125 KDA PROTEIN.
OS   Alfalfa mosaic virus (strain 425 / isolate Leiden).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC   Alfamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83143345; PubMed=6298738;
RA   Cornelissen B.J.C., Brederode F.T., Moormann R.J.M., Bol J.F.;
RT   "Complete nucleotide sequence of alfalfa mosaic virus RNA 1.";
RL   Nucleic Acids Res. 11:1253-1265(1983).
CC   -!- ---------------------------------------------------------------------
CC   -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   -!- between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   -!- the European Bioinformatics Institute. There are no restrictions on its
CC   -!- use by non-profit institutions as long as its content is in no way
CC   -!- modified and this statement is not removed. Usage by and for commercial
CC   -!- entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   -!- or send an email to license@isb-sib.ch).
CC   -!- ---------------------------------------------------------------------
DR   EMBL; L00163; AAA46289.1; -.
DR   PIR; A04197; WMFM12.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   1126 AA;  125827 MW;  BF5A8019B47D4CBF CRC64;


alignment_scores:
            Quality:  329.50              Length:   1884
              Ratio:  0.504                 Gaps:     63
 Percent Similarity:  34.713   Percent Identity: 16.879

alignment_block:
 US-09-301-906-3 x V125_AMVLE   ..

 Align seg 1/1  to: V125_AMVLE  from: 1  to: 1126

    1303 ACGAAATTGTGCGATATGTTTTCTCAA.................AGGGA 1334
         |||   :::      :::||||||:::                 :::||
      40 ThrAlaIleGlyLysValPheSerGluAlaGlyArgAlaTyrAlaGlnAs 56

    1335 CGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTT..... 1377
         ||||:::           ||||||   :::   :::|||:::
      56 pAlaLeu............ProSerAspLysGlyGluValLeuLysIleS 69

    1378 ........CTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCA 1419
                 |||         :::::::::   ||||||      ||||||
      69 erPheSerLeuAspAlaThrGlnGlnAsnIleLeuArgAlaAsnPhePro 85

    1420 GAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGC 1469
                       ||||||:::|||      |||||||||   |||'|
      86 GlyArgArgThrValPheSerAsnSerSerSerSerSerHisCysPheAl 102

    1470 TAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTA 1519
         |   |||   |||         :::   :::      |||      |||
     102 aAlaAlaHisArgLeuLeuGluThrAspPheValTyrArgCysPheGlyA 119

    1520 ATGTG......TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTC 1563
         ||:::         ::::::|||:::||||||:::|||:::   |||:::
     119 snThrValAspSerIleIleAspLeuGlyGlyAsnPheValSerHisMet 135

    1564 AAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAA 1613
         |||:::   :::   |||   |||   |||   |||:::||||||:::::
     136 LysValLysArgHisAsnValHisCysCysCysProIleLeuAspAlaAr 152

    1614 AGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGG 1663
         :|||       |||       ::::::||||||   |||:::
     152 gAspGlyAlaArgLeuThrGluArgIleLeuSerLeuLys.......... 165

    1664 GAGATTCGTACGTGTCCAGT......GACCTTCTAACTGAAGCGGCTTCA 1707
               |||||||||   :::      :::::::::   ||||||
     166 .....SerTyrValArgLysHisProGluIleValGlyGluAla...... 178

    1708 AAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGC 1757
                 :::||||||            |||:::|||:::   |||||
     179 .........AspTyrCysMetAspThrPheGlnLysCysSerArgArgAl 192

    1758 CGATGCGGGTTTTATGGTG...GATGTGTACGATATATCCCCGCAGCAGG 1804
         ||||   :::|||   :::   ::::::   |||::::::      ::::
     192 aAspTyrAlaPheAlaIleHisSerThrSerAspLeuAspValGlyGluL 209

    1805 TAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTT 1854
         ::|||   :::::::|||:::||||||::::::   |||      ::::::
     209 euAlaCysSerLeuAspGlnLysGlyValMetLysPheIleCysThrMet 225

    1855 ATGTTCCCCGTGGAGTTGTTGTACGGTAAC...GGTGAAGTTTACTTGGA 1901
         |||       ::::::::::|||      |||   ||||||:::
     226 MetValAspAlaAspMetLeuIleHisAsnGluGlyGluIleProAsnPh 242

    1902 AGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTAC.... 1947
          ::::::        :::   |||:::   |||   :::   :::
     242 eAsnValArgTrpGluIleAspArgLysLysAspLeuIleHisPheAspP 259

    1948 .............AATGTTGGTCAGTGTGGTGAGATGTATGAACATTCC 1983
```

```
                  |||:::|||                |||:::|||
 259 heIleAspGluProAsnLeuGly.............TyrSerHisArg 270

1984 TTCTCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGCACTTCGTC 2033
     ||||||   ::::::   :::::::|||:::
 271 PheSerLeuLeuLysHisTyrLeuThrTyr..................... 280

2034 CGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATC 2083
       |||:::      |||   :::   :::||||||
 281 ....AsnAlaValAspLeuGlyHisAlaAlaTyrArg.............. 291

2084 ATCTCACTATGTGTAGGGCTCAGAAGTCACCTGGAACT.........GAG 2124
            :::   |||   |||         |||:::         :::
 292 ........IleGluArgLysGlnAspPheGlyGlyValMetValIleAsp 305

2125 GTTACGTATAGGTCGTTG..............GTCCCGTCGTTCGTGGG 2159
     :::||||||   ||||||                :::|||         ||
 306 LeuThrTyr...SerLeuGlyPheValProLysMetProHisSerAsnGl 321

2160 CAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG....... 2202
     |:::|||   ::::::        |||   |||:::   |||
 321 yArgSerCysAlaTrpTyrAsnArgValLysGlyGlnMetValValHisT 338

2203 .........................TCCTTTAAGACAATAGTCCTCGAT 2226
                               |||:::::::|||   |||
 338 hrValAsnGluGlyTyrTyrHisHisSerTyrGlnThrAlaValArgArg 354

2227 TCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGAC 2276
     :::   :::||||||:::                 ::::::   ::::::
 355 LysValLeuValAspLysLysValLeuThrArgValThrGluValAlaPh 371

2277 ATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAA 2326
       :::   |||          |||   :::|||:::::::|||
 371 eArgGlnPheArgProAsnAlaAspAlaHisSerAlaIleGlnSerIleA 388

2327 AGACC.....................CATGTCATTACAGGGAGTCGCGTT 2355
       |||                     |||:::|||   |||:::   :::
 388 laThrMetLeuSerSerSerThrAsnHisThrIleIleGlyGlyValThr 404

2356 GTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGT 2405
     :::   |||         :::||||||||||||         ::::::::
 405 LeuIleSerGlyLysProLeuSerProAspAspTyrIleProValAlaTh 421

2406 CGCTGTTATGGCTCAGGCGATTAAG.......................... 2430
     :::::::      ::::::   |||
 421 rThrIleTyrTyrArgValLysLysLeuTyrAsnAlaIleProGluMetL 438

2431 ...........GATAGGGCGAAGAGTATTCGCTCCTATAACTTTATAAAA 2469
                |||:::::::::   :::   :::         :::|||
 438 euSerLeuLeuAspLysGlyGluArgLeuSerThrAspAlaValLeuLys 454

2470 GCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGACCGT 2519
     :::|||||||||   :::
 455 GlySerGluGlyProMet.....................TrpTyrSe 463

2520 AGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGC 2569
      |||      |||   :::|||::::::         |||:::   |||
 463 rGlyProThrPheLeuSerAlaLeuAsp.........LysValAsnValP 477

2570 ACGATAACTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTC 2619
          :::|||   |||   :::::|||:::|||:::|||:::
 477 roGlyAspPhe...ValAlaLysAlaLeuLeuSerLeuProLys...... 490

2620 ATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGC 2669

 490 ................................................... 490

2670 TTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCT 2719

 490 ................................................... 490

2720 TCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATC 2769
          |||   ||||||:::            :::||||||||||:::

 491 .....ArgAspLeuLysSer.............LeuPheSerArgSer... 500

2770 GTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAA 2819
                                                |||:::   :::::
 501 ................................AlaThrSerHisSerGl 506

2820 ACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACG 2869
     :      |||:::   ::::::
 506 uArgThrProValArgAspGlu............................ 513

2870 TTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCG 2919
       |||   :::::::|||
 514 ..SerProIleArgCys................................... 518

2920 ACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGAC 2969

 518 ................................................... 518

2970 GAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCT 3019
                                          |||   ::::::||||
 519 ................................ThrAspGlyValPheT 524

3020 ACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACA 3069
     ||
 524 yr................................................. 524

3070 GAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGC 3119
                        |||:::   :::|||
 525 ..................ProIleArgMetLeu.................. 529

3120 TGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCT 3169
           |||   :::::::::   |||       ::::::::::|||::::
 530 ....LeuLysCysLeuGlySerAspLys......PheGluSerValThrI 543

3170 TAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATA 3219
     :::::|||
 543 leThrAsp............................................ 545

3220 TTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGG 3269

 545 ................................................... 545

3270 AATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAA 3319

 545 ................................................... 545

3320 GCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGC 3369
                                        |||:::   |||::::::
 546 ............................ProArgSerAsnThrGluThr 552

3370 ACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGT 3419
     |||||||:::                :::   ::::::|||::::::   ::
 553 ThrValAsp..............LeuTyrGlnSerPheGlnLysLysIl 564

3420 GACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGC 3469
     :
 564 e.................................................. 564

3470 TATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATT 3519

 564 ................................................... 564

3520 TGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGAC 3569
                          :::   :::|||   |||   ||||||
 565 ....................GluThrValPheSerPheIleLeuGlyLy 574

3570 ATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCT 3619
      :::   |||
 574 sIleAspGlyPro...................................... 578

3620 CCAATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTC 3669

 578 ................................................... 578
```

```
3670 AGCTCATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGG 3719

 578 ..................................................  578

3720 CTTGCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGG 3769
                                                 |||:::   :
 579 .................................SerProLeuIleS 583

3770 GGAGGAATGTGTACGATGAGACGCTTAGG......TATTACCGGGACTTT 3813
     ::      ||||||   ::::::|||:::      ||||||   ::::::
 583 erAspProValTyrPheGlnSerLeuGluAspValTyrTyrAlaGluTrp 599

3814 GACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCC 3863

 599 ..................................................  599

3864 CGGTGACGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTG 3913
               :::::::|||::::::   ::::::|||:::|||
 600 ..........HisGlnGlyAsnAlaIleAspAlaSerAsnTyr.......  610

3914 TCACAAATGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGT 3963

 610 ..................................................  610

3964 GAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGA 4013

 610 ..................................................  610

4014 AGAGGTTGCCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGG 4063

 610 ..................................................  610

4064 GTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGT 4113
                                        ::::::
 611 .....................................AlaArgThrLeu  614

4114 GTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGT 4163
     :::||||||:::      |||:::::::::::::      |||:::
 615 LeuAspAspIleArgLysGlnLysGluGluSerLeuLysAlaLysAlaLy 631

4164 TGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGC 4213
      ||||||:::          |||   ::::::::::::
 631 sGluValGluAspAlaGlnLysLeuAsnArgAlaIleLeu..........  644

4214 CGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCG 4263

 644 ..................................................  644

4264 GTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCC 4313

 644 ..................................................  644

4314 ACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAG 4363
                                     :::||||||||||      |
 645 ............................GlnValHisAlaTyrLeuG  651

4364 AAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATC 4413
     ||:::      :::
 651 luAlaHisProAsp.....................................  655

4414 GTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAAT 4463
                                           |||   ||||||
 656 ..................................GlyGlyLysIleGl  660

4464 TGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAG 4513
           |||   :::      :::|||:::      ::::::   :
 660 uGlyLeuGlyLeuSerSerGlnPheIleAlaLysIleProGluLeuAlaI 677

4514 TCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGT 4563
     ::      ||||||                ::::::::::::::|||
 677 leProThrProLysProLeuProGluPheGluLysAsnAlaGluThrGly 693

4564 AAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGT 4613
     :::   |||                              :::::::::
 694 GluIleLeu............................ArgIleAsnPr 700

4614 GTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTG 4663
      ::::::::::|||   :::                :::|||:::   ::::
 700 oHisSerAspAlaIleLeu..............GluAlaIleAspTyrL 712

4664 TCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAG 4713
     ::   :::   :::::::::|||   :::   :::::|||||||||
 712 euLysSerThrSerAlaAsnSerIleIleThrLeuAsnLysLeuGly... 727

4714 GGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGC 4763

 727 ..................................................  727

4764 CCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACT 4813
                                                 |||||||
 728 .........................................AspHisC  730

4814 GTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGAT 4863
     ||      |||          |||      |||   :::   :::|||
 730 ys......GlnTrpThrThrLysGlyLeuAspValValTrpAlaGlyAsp 744

4864 GACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGT 4913
     |||:::
 745 AspLys.............................................  746

4914 GGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGG 4963

 746 ..................................................  746

4964 TCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAA 5013

 746 ..................................................  746

5014 AGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTT 5063

 746 ..................................................  746

5064 GACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGT 5113
                     ||||||:::|||
 747 ...............ArgArgAlaPhe.....................  750

5114 TAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAA 5163
                                        :::|||::::::::::
 751 ..................................IleProLysLysAsn  755

5164 CCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAA 5213
                     ::::::|||:::             :::      ::
 756 ThrTrpValGlyProThrAlaArgSerTyrProLeuAlaLysTyrGluAr 772

5214 GGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACT 5263
     :|||:::|||
 772 gAlaMetSer.....................................  775

5264 CGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCT 5313
               ::::::   ||||||              ||||||
 776 ...........LysAspGlyTyrValThr...........LeuArgTrp 784

5314 GATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGA 5363
     |||:::   |||:::      :::         :::         :::||
 785 AspGlyGluValLeuAspAlaAsnCysValArgSerLeuSerGlnTyrGl 801

5364 AGTGGTTAAAACCGGG..............ACCTCTGAAGATGCCGTCG 5398
     |:::|||   :::                  :::|||   :::|||::::
 801 uIleValPheValAspGlnSerCysValPheAlaSerAlaGluAlaIleI 818

5399 TGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCG 5448
     ::      |||:::                              :::|||
 818 leProSerLeuGlu............................LysAla 824

5449 TTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTG.. 5496
```

```
     |||   :::      ::::::::::||||||   ::::::||||||
  825 LeuGlyLeuGluAlaHisPheSerValThrIleValAspGlyValAlaGl 841

 5497 .......AAAGTACCTAAT......CAAGTTTATGAATCACTACCGGGCT 5533
             |||:::   |||      |||:::   :::|||
  841 yCysGlyLysThrThrAsnIleLysGlnIleAlaArgSer..........  854

 5534 TTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGAC 5583

  854 .................................................. 854

 5584 GGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTAT 5633

  854 .................................................. 854

 5634 CTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATC 5683
             :::|||:::|||||||||   :::      |||   :::|||
  855 .......SerGlyArgAspValAspLeuIle......LeuThrSerAsnA 867

 5684 TGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATA 5733
             :::|||::::::              |||:::      |||
  867 rgSerSerAlaAspGluLeuLysGluThrIleAspCysSerProLeuThr 883

 5734 GGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTC 5783
                  :::      |||
  884 LysLeuHisTyrIleArgThrCys..........................  891

 5784 CTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGA 5833

  891 .................................................. 891

 5834 CGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACG 5883
             |||::::::
  892 ........AspSerTyr.................................. 894

 5884 GCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGA 5933

  894 .................................................. 894

 5934 TCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGG 5983
                     |||   ||||||::::::|||::::::
  895 ................LeuMetSerAlaSerAlaValLysAla....... 903

 5984 TAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTG 6033
                               :::|||::::::   ||||||
  904 ........................GlnArgLeuIlePheAspGluCys 911

 6034 CACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGC 6083
        :::::::|||   ||||||:::      ::::::      :::|||::
  912 PheLeuGlnHisAlaGlyLeuValTyrAlaAlaAlaThrLeuAlaGlyCy 928

 6084 GTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACA 6133
      :||||||   :::   |||||||||   :::||||||||||||::::::|
  928 sSerGluValIleGlyPheGlyAspThrGluGlnIleProPheValSerA 945

 6134 GGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAA 6183
      ||:::      |||   :::           ::::::   |||   |||
  945 rgAsnProSerPheValPheArgHisHisLysLeuThrGlyLysValGlu 961

 6184 AGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTA 6233
           :::   |||         :::|||   |||   |||:::   ||
  962 ArgLysLeuIleThr.........TrpArgSerProAlaAspAlaThrTy 975

 6234 CTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAA 6283
      |   |||                                     ||||
  975 rCysLeu.....................................GluL 979

 6284 AGGTCGTTAGCGGTAAGGACAAACCA.................GTAGTA 6315
      ||           :::   ||||||                 |||:::
  979 ysTyrPheTyrLysAsnLysLysProValLysThrAsnSerArgValLeu 995

 6316 AGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAAT 6365
      ||||||::::::      ||||||::::::      :::|||   :::
```

```
  996 ArgSerIleGluValValProIleAsnSerProValSerValGluArgAs 1012

 6366 AAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGA 6415
       :::      :::||||||||||   ||||||   ||||||:::   :::|
 1012 nThrAsnAlaLeuTyrLeuCysHisThrGlnAlaGluLysAlaValLeuL 1029

 6416 AGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCAT 6465
      ||   :::         ||||||   ::::::   ::::::|||:::|||
 1029 ysAlaGlnThrHisLeuLysGlyCysAspAsn...IlePheThrThrHis 1044

 6466 GAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAA 6515
      |||||||||||||||||||||:::::::|||   :::   |||      ::
 1045 GluAlaGlnGlyLysThrPheAspAsnValTyrPheCysArgLeuThrAr 1061

 6516 AGCCGATGACTCCCTATTCACTAAACAACCGCATATA.............. 6552
      ::::::::  ||||||   |||   :::      |||
 1061 gThrSerThrSerLeuAlaThrGlyArgAspProIleAsnGlyProCysA 1078

 6553 .....CTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCT 6597
           ||||||:::|||||||||||   :::::::::   |||   :::
 1078 snGlyLeuValAlaLeuSerArgHisLysLysThrPheLysTyrPheThr 1094

 6598 CTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTC 6647
      ::::::         ||||||   :::   :::         ||||||::
 1095 IleAlaHisAspSerAspAspValIleTyrAsnAlaCysArgAspAlaGl 1111

 6648 ACCTCAATCAGTATCCGACGCTTTGCTT..............CACACGT 6682
      :      ::::::::::|||::::::|||               |||:::|
 1111 y......AsnThrAspAspSerIleLeuAlaArgSerTyrAsnHisAsnP 1126

 6683 TC 6684
      ||
 1126 he 1126
```

```
seq_name: SwissProt_39:RRPO_TOMS1

seq_documentation_block:
ID   RRPO_TOMS1     STANDARD;      PRT;  1616 AA.
AC   Q9YKD6; Q9WJ37;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain S-1) (ToMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT   "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
RT   virus.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ132845; CAB36997.1; -.
DR   EMBL; AJ132845; CAB36998.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
```

```
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183542 MW;  5DBBFB2FADCC5C0C CRC64;


alignment_scores:
            Quality:  313.00              Length:   1879
              Ratio:  0.455                 Gaps:     63
 Percent Similarity:  36.615   Percent Identity: 16.871

alignment_block:
 US-09-301-906-3 x RRPO_TOMS1   ..

 Align seq 1/1  to: RRPO_TOMS1  from: 1  to: 1616

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||      :::   |||:::::::::::::::::|||   ::
       2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrIleArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::  :::|||  ::::::  |||                 ||||||
      18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::  |||  |||:::  ||||||           ::::::|||
      35 laVal...AspGluPheAsnAlaArgAsp...........ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::  :::  :::      ::::::  ::::::
      47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
          :::  :::|||||::::::|||  |||  ::::::  ::::::
      63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||  :::|||:::::::::||||||      ::::::  ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
           :::         :::         :::|||||||||||||:::|||::
      97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

    1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
         :  |||:::      |||::::::      |||  |||  |||   |
     113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

    1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG...........ATC 1641
         ||||||||:::|||:::  |||:::  ::::::
     130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

    1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCCAGT..... 1683
         |||:::||||||           |||:::::::::|||  :::
     147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

    1684 ............GACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTT 1720
                     |||     :::|||     ::::::  |||   :
     163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

    1721 AC.............TGTAGTCGAGAATCGCAGAACTGCGATTCTAGA 1755
         ::             |||:::::::::::::::        :::|||
     180 isAspThrPheGlnThrCysGlyHisSerGlnGluCysTyrThrGlyArg 196

    1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
         :::  |||      :::  ::::::|||||||||      ::::::
     197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

    1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
          :::  |||:::  :::|||::::::  |||        |||:::
```

```
     213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

    1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
           |||      :::||||||      ::::::::::|||   |||:::|||
     230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

    1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
         :::::::::      :::|||:::||||||  |||::::::::::  ::
     247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

    1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
         ::::  :::  :::  |||:::||||||:::||||||:::      :
     263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280

    2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
         ::         :::|||      :::|||:::  :::|||:::
     280 yrValCysLysThrTyrProAlaSerAsnArgGluValTyrMetLys 296

    2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
         |||:::
     297 GluPhe..................................... 298

    2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

     298 ..................................................... 298

    2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

     298 ..................................................... 298

    2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                          :::|||  |||:::  ::
     299 .........................LeuValThrArgValAsnTh 305

    2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
         ::::        ::::::  |||  ||||||
     305 rTrpPheCysLysPheSerArgIleAspThrPhe............... 316

    2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                                  :::
     317 ....................................LeuLeuTyr 319

    2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
         :::  |||:::|||:::  ||||||
     320 LysGlyValAlaHisLysGlyValAsp..................... 328

    2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446

     328 ..................................................... 328

    2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
              |||   :::|||   ||||||  |||  ::
     329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341

    2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
                   :::  :::|||::::::  |||  |||:::  :::
     342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

    2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
          :::      ::::::  |||:::      |||      ::::::  :
     355 uLeuGluAspSerSerSerValAsnTyrTrpPheProLysMetArgAspM 372

    2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
         :::::  :::|||
     372 etValIleValPro.............................. 376

    2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

     377 ..............................................Le 377

    2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
          ||||||||||||  ||||||              |||::::::|||
     377 uPheAspIleSerLeuGluThr............SerLysArgThrArgL 390
```

```
2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
     :::|||:::          ::::::::::|||   :::      |||::: 
 390 ysGluValLeu.........ValLeuLysAspPheValTyrThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   ::::::|||   :::::::|||   |||      |||   ::::: 
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu......ThrTyrSerAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::      |||:::||| 
 418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
          ::::::::::|||      |||          :::|||:::||| 
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||                                   :::   || 
 446 SerLeu.................................LeuGlnSe 450

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
     |:::      |||      |||             ::::::   :::|||: 
 450 rLeuSerMetThrPhePheLeuHisIleLysLeuAlaValLeuLysAspA 467

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
     ::   ||| 
 467 spLeuLeu..................................... 469

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
                              :::|||||   :::   ||| 
 470 ........................IleSerLysPheAlaLeuGlyPr 477

3144 GAAA...........GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
      |||           :::   |||:::|||:::|||||| 
 477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu..... 490

3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231

 490 ..................................................  490

3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281

 490 ..................................................  490

3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331

 490 ..................................................  490

3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381

 490 ..................................................  490

3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
                         :::|||:::::::|||            :: 
 491 .....................AlaPheGlyAsnAlaPheProSerIleLy 500

3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
     :      :::   :::|||:::          -                 . 
 500 sGluArgLeuIleAsnArgLys..... 507

3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
                               ::::::   |||      :::||| 
 508 .........................LeuIleLysIleThrGluAsnAla 515

3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
     |||
 516 LeuGlu..................................... 517

3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631

 518 .....................................................I 518
```

```
3632 GCGCTGTTCCCGAG......GTTGTTTGGCGAGGGAAGAGTGTCAGCTCA 3675
          ||||||:::       |||::::::::::   :::   ||||||::: 
 518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534

3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
          :::   :::   ::: 
 535 TyrLysMetSerValAspMet..... 541

3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
                                     |||   ::::::   |||: 
 542 ..............................ProValLeuAspIleArgL 548

3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
     ::       :::||||||   :::   ||| 
 548 ysLysMetGluGluThrGluGluMetTyr..... 557

3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875

 557 ..................................................  557

3876 CGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGC 3925
                    ::::::::::|||:::   :::|||:::   ||| 
 558 .............AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570

3926 GCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACC 3975
                  :::::::::   ::::::::: 
 570 spLysPheAspValAspValPheSerGlnMetCys..... 581

3976 CATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCC 4025
               :::|||:::                        ||||||   || 
 582 .........GlnSerLeu.....................GluValAspPr 588

4026 GTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG 4075
     |   :::|||                                       | 
 588 oMetThrAla.....................................A 592

4076 CTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTC 4125
     ||   :::::::::   |||   |||   ::::::|||:::      ::: 
 592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuThr 608

4126 TTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGG 4175
     |||:::|||   ::::::::::::::|||      ::: 
 609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu..... 621

4176 CAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAA 4225
                                                 |||::: 
 622 .........................................GlnAspS 624

4226 GGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGC 4275
          :::|||:::||||||
 624 erGluLysAlaSerAspGly..... 630

4276 GACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATAT 4325
                       :::   :::|||:::|||   ::::::::: 
 631 ..................AlaLeuValValThrSerArgAspValGluGl 641

4326 TCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAG 4375
      |||::::::      |||                 . ::::::   | 
 641 uProSerIleLysGlySer..................MetAlaArgGlyG 652

4376 AAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425
     ||:::   ||||||   :::||||||                  ::: 
 652 luLeuGlnLeuAlaGlyLeuSerGly.................AspVal 662

4426 CCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAA 4475
     |||       |||   ||| 
 663 ProGluSerSerTyrThr..... 668

4476 GGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCA 4525
                              ::::::   ::::::|||:::   : 
 669 .........................ArgSerGluGluIleGluSerLeuG 677

4526 AACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGC 4575
```

```
     ::|||      ::::::      ::::::         ||||||:::|||
 677 luGlnPheHisMetAlaThrAlaSerSerLeuIleHisLysGlnMetCys 693

4576 ATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGC 4625
       :::                                ||||||:::
 694 SerIle............................ValTyrThrGlyPr 700

4626 GACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGA 4675
           :::   ::::::::::|||         |||:::|||:::|||
 700 oLeuLysValGlnGlnMetLysAsn.........PheIleAspSerLeuV 714

4676 AAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTAT 4725
       :::      ::::::
 714 alAlaSerLeuSerAla.................................. 719

4726 AATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATAT 4775
                                        |||:::::::::::
 720 ....................................AlaValSerAsnLe 724

4776 CTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGA 4825
     ::::      :::|||      :::::::::|||      |||      :
 724 uValLysIleLeuLysAspThrAlaAlaIleAsp......LeuGluThrA 739

4826 AGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGC 4875
     ::   |||:::|||
 739 rgGlnLysPheGly..................................... 743

4876 TATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAA 4925

 743 ..................................................... 743

4926 CTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAG 4975
                                        |||:::::::||||
 744 ....................................ValLeuAspValA 748

4976 CTTCGGGTGACTATTTTCTTATGCCTTGC.................GGT 5007
     |||||      :::::::::   |||                     |||
 748 laSerLysArgTrpLeuValLysProSerAlaLysAsnHisAlaTrpGly 764

5008 TTTCAAAGGACGCACTTGCATTCAGTAAAC............TCCATCGA 5045
          :::||||||   ::::::   :::              :::   ||
 765 ValValGluThrHisAlaArgLysTyrHisValAlaLeuLeuGluHisAs 781

5046 CGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTG 5095
     ||||                                          ||||||:
 781 pGlu.....................................PheGlyI 785

5096 TAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCA 5145
     ::
 785 le................................................ 785

5146 GGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCAC 5195

 785 ..................................................... 785

5196 ACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCA 5245

 785 ..................................................... 785

5246 AGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAG 5295
                       |||   |||:::
 786 ...................IleThrCysAsp..................... 789

5296 TGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTA 5345
        :::|||   |||              :::|||::::::|||
 790 ...AsnTrpArgArg..............ValAlaValSerPro..... 798

5346 CACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCG 5395

 798 ..................................................... 798

5396 TCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGG 5445
          |||   :::   |||      ::::::      ||||||   |||
```

```
 799 .....GluSerValValTyrSerAspMetAlaLysLeuArgThrLeuArg 813

5446 GCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCT 5495
        ||||||          :::      |||::::::|||:::   |||||
 814 ArgLeuLeuLysAspGlyGluProHisValSerSerAlaLysValValLe 830

5496 GAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACA 5545
     |
 830 u..................................................... 830

5546 AGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTG 5595

 830 ..................................................... 830

5596 AGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGG 5645

 830 ..................................................... 830

5646 CAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCG 5695
           ||||||:::|||
 831 .......ValAspGlyVal.............................. 834

5696 ACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAA 5745

 834 ..................................................... 834

5746 GTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAA 5795

 834 ..................................................... 834

5796 GTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACG 5845
                    ||||||   ||||||
 835 ................ProGlyCysGlyLys.................. 839

5846 AATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA 5895
                              |||      |||   ::::::|||:::
 840 ......................ThrLysGluIleLeuSerArgValAsn 848

5896 AGTTCTGAGGACATAAATATGGCG.....................GTGAA 5924
       :::||||||:::   ::::::                     :::
 849 PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetIl 865

5925 GAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACT 5974
      :::|||      |||   :::|||:::      ||||||      |||
 865 eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882

5975 CCAGGGTGGTA...AACTTTATCGTC..................AGGGGA 6003
       |||:::|||   :::|||::::::                   |||
 882 alArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898

6004 ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTT 6053
     ::::::||||||:::::::::||||||      |||:::|||   ||
 899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyCy 915

6054 ACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTG 6103
      :::::::::   |||      :::   ::::::   :::      :::|
 915 sValAsnPheLeuValGluMetSerLeuSerAspIleAlaTyrValTyrG 932

6104 GAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATG 6153
     |||||   :::||||||||||:::||||||||         |||
 932 lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948

6154 GATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAA 6203
       :::          :::   ::::::|||
 949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965

6204 ATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA....... 6246
       :::   |||||||||   ||||||   :::::::|||::::::
 965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982

6247 ..........ATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAG 6285
                |||   :::   :::   :::|||         ::::::
 982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998
```

```
    6286 GTCGTTAGCGGT......AAGGACAAACCAGTAGTAAGATCGCTGTCCAA 6329
         :::|||||||||       :::   :::||||||   :::   |||:::
     999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl 1015

    6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT 6379
         :::
    1015 yLys.....................................I 1017

    6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
           |||   :::||||||          ||||||   |||:::::::|||
    1017 leLeuThrPheThrGln.........SerAsp...LysGluAlaLeuLeu 1029

    6430 GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA 6479
         ::::::|||      |||   |||   |||||||||||:::||||||::
    1030 SerArgGly...TyrThrAspValHisThrValHisGluValGlnGlyGl 1045

    6480 AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC 6529
         :|||::::::||||||   |||   |||          :::      |||:
    1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062

    6530 TATTCACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
         ::::::::::   :::||||||:::||||||:::|||||||||||||||
    1062 leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078

    6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
         :::||||||   |||   :::                |||   |||   ::
    1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleIl 1095

    6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
         :   |||      :::|||::::::|||:::   |||
    1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq_name: SwissProt_39:RRPO_TOML

seq_documentation_block:
ID   RRPO_TOML      STANDARD;      PRT;  1616 AA.
AC   P03587; O41352;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain L) (ToMV) (TMV strain tomato).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85157522; PubMed=6549393;
RA   Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
RA   Okada Y.;
RT   "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT   genome and comparison with the common strain genome.";
RL   J. Biochem. 96:1915-1923(1984).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02144; CAA26085.1; -.
DR   EMBL; X02144; CAA26082.1; -.
DR   PIR; A04195; WMTM8T.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183564 MW;  A8EC8929B5CF7CAF CRC64;


alignment_scores:
            Quality:  309.00              Length:   1879
              Ratio:  0.445                 Gaps:     64
 Percent Similarity:  36.935   Percent Identity: 16.764

alignment_block:
 US-09-301-906-3 x RRPO_TOML   ..

 Align seg 1/1  to: RRPO_TOML  from: 1  to: 1616

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||       :::   |||:::::::::::::::::::   ::
       2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::   :::|||   ::::::   |||                 ||||||
      18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::   |||   |||:::   ||||||            :::::|||
      35 laVal...AspGluPheAsnAlaArgAsp............ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::   :::   :::       ::::::  ::::::
      47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
          :::   :::||||||:::::::|||   |||   ::::::   ::::::
      63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||   :::|||:::::::::||||||      ::::::   ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
         :::          :::       :::||||||||||||:::|||::
      97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

    1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
         :   |||:::      |||::::::      |||   |||   |||   |
     113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

    1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG............ATC 1641
         ||||||||:::|||:::   |||:::   ::::::
     130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

    1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCC........ 1680
         |||:::||||||          |||:::::::::|||
     147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

    1681 ......................AGTGACCTTCTAACTGAAGCGGCTTCAA 1708
                                :::::::::   :::|||::::::
     163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

    1709 AGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC...GATTCTAGA 1755
           ::::::          |||:::||||||:::|||      :::|||
     180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196

    1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
         :::   |||      :::   ::::::||||||||||      ::::::
     197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213
```

```
1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
      :::   |||:::   :::|||:::::   |||         |||:::
 213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
       |||      :::||||||      :::::::::|||   |||:::|||
 230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
     :::::::::      :::|||:::||||||   |||::::::::::   ::
 247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
     ::::   :::   :::   |||:::||||||:::|||||||:::      :
 263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280

2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
     ::         :::|||      :::|||:::   :::|||:::
 280 yrValCysLysThrTyrPheProAlaSerAsnArgGluValTyrMetLys 296

2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
     |||:::
 297 GluPhe..................................................... 298

2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

 298 ..................................................... 298

2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

 298 ..................................................... 298

2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                      :::|||   |||:::   ::
 299 ..........................LeuValThrArgValAsnTh 305

2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
     ::::         ::::::   |||   ||||||
 305 rTrpPheCysLysPheSerArgIleAspThrPhe................ 316

2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                                  :::
 317 .....................................LeuLeuTyr 319

2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
     :::   |||:::|||:::   ||||||
 320 LysGlyValAlaHisLysGlyValAsp..................... 328

2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446

 328 ..................................................... 328

2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
          |||   :::|||   ||||||   |||   :::
 329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis...... 341

2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
                :::   :::|||::::::   |||   |||:::   :::
 342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
      :::      ::::::   |||:::      |||      ::::::   :
 355 uLeuGluAspSerSerSerValAsnTyrTrpPheProLysMetArgAspM 372

2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
     :::::   :::|||
 372 etValIleValPro..................................... 376

2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

 377 ..................................................Le 377

2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746

      ||||||||||||   ||||||              |||::::::|||
 377 uPheAspIleSerLeuGluThr............SerLysArgThrArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
       :::|||:::          :::   :::|||   :::      |||:::
 390 ysGluValLeu.........ValSerLysAspPheValTyrThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   ::::::|||   :::::::|||   |||      |||   :::::
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu......ThrTyrSerAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::      |||:::|||
 418 nValLeuSerPheValGluSerIleArgSerArgVal.............. 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
           :::::::::|||      |||           :::|||:::|||
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||:::         :::   |||
 446 SerLeuLeuGlnSerLeuSerMetThr....................... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
      ||||||
 455 .PhePhe..................................... 456

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
                                         :::::::::   :::
 457 ..............................LeuHisThrLysLeu 461

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
     ::::::|||:::   :::   :::   :::||||||   :::   |||
 462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477

3144 GAAA............GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
      |||           :::   |||:::|||:::||||||
 477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu.......... 490

3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231

 490 ..................................................... 490

3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281

 490 ..................................................... 490

3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331

 490 ..................................................... 490

3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381

 490 ..................................................... 490

3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
                              :::|||::::::|||            ::
 491 ....................AlaPheGlyAsnAlaPheProSerIleLy 500

3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
     :     :::   :::|||:::
 500 sGluArgLeuIleAsnArgLys............................ 507

3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
                                   ::::::   |||      :::|||
 508 ........................LeuIleLysIleThrGluAsnAla 515

3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
     |||
 516 LeuGlu.......................................... 517

3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631
```

```
 518 ..............................................I 518

3632 GCGCTGTTCCCGAG......GTTGTTTGGCGAGGGAAGAGTGTCAGCTCA 3675
       ||||||::      |||:::::::::   :::   ||||||:::
 518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534

3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
         :::   :::   :::
 535 TyrLysMetSerValAspMet.......................... 541

3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
                              |||   ::::::   |||:
 542 ............................ProValLeuAspIleArgL 548

3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
     ::      :::|||||||   :::   |||
 548 ysLysMetGluGluThrGluGluMetTyr..................... 557

3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875

 557 ..................................................... 557

3876 CGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGC 3925
                 :::::::::|||:::   :::|||:::   |||
 558 .............AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570

3926 GCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACC 3975
                :::::::::   :::::::::
 570 spLysPheAspValAspValPheSerGlnMetCys.............. 581

3976 CATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCC 4025
             :::|||:::                    ||||||   ||
 582 .........GlnSerLeu.....................GluValAspPr 588

4026 GTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG 4075
     |   :::|||                                    |
 588 oMetThrAla.....................................A 592

4076 CTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTC 4125
     ||   :::::::::   |||   |||   :::::::|||:::      :::
 592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuThr 608

4126 TTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGG 4175
     |||:::|||   ::::::::::::::::|||      :::
 609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu........... 621

4176 CAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAA 4225
                                                 |||:::
 622 .............................................GlnAspS 624

4226 GGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGC 4275
       :::|||:::||||||
 624 erGluLysAlaSerAspGly............................. 630

4276 GACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATAT 4325
                    :::   :::|||:::|||   :::::::::
 631 ..................AlaLeuValValThrSerArgAspValGluGl 641

4326 TCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAG 4375
      |||::::::      |||                    ::::::   |
 641 uProSerIleLysGlySer..................MetAlaArgGlyG 652

4376 AAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425
     ||:::   ||||||   :::||||||                  :::
 652 luLeuGlnLeuAlaGlyLeuSerGly..................AspVal 662

4426 CCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAA 4475
     |||      |||   |||
 663 ProGluSerSerTyrThr.............................. 668

4476 GGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCA 4525
                              ::::::   ::::::|||:::   :
 669 .......................ArgSerGluGluIleGluSerLeuG 677

4526 AACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGC 4575
     ::|||      ::::::      ::::::         ||||||:::|||
 677 luGlnPheHisMetAlaThrAlaSerSerLeuIleHisLysGlnMetCys 693

4576 ATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGC 4625
        :::                                 |||||||:::
 694 SerIle.............................ValTyrThrGlyPr 700

4626 GACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGA 4675
           :::   :::::::::|||        |||:::|||:::|||
 700 oLeuLysValGlnGlnMetLysAsn.........PheIleAspSerLeuV 714

4676 AAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTAT 4725
       :::      ::::::
 714 alAlaSerLeuSerAla.................................. 719

4726 AATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATAT 4775
                                        |||::::::::::::
 720 ..................................AlaValSerAsnLe 724

4776 CTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGA 4825
     ::::      :::|||      :::::::::|||      |||      :
 724 uValLysIleLeuLysAspThrAlaAlaIleAsp......LeuGluThrA 739

4826 AGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGC 4875
     ::   |||:::|||
 739 rgGlnLysPheGly................................... 743

4876 TATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAA 4925

 743 ..................................................... 743

4926 CTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAG 4975
                                        |||::::::||||
 744 ..................................ValLeuAspValA 748

4976 CTTCGGGTGACTATTTTCTTATGCCTTGC..................GGT 5007
     |||||      :::::::::   |||                     |||
 748 laSerLysArgTrpLeuValLysProSerAlaLysAsnHisAlaTrpGly 764

5008 TTTCAAAGGACGCACTTGCATTCAGTAAAC............TCCATCGA 5045
          :::||||||   ::::::   :::            :::   ||
 765 ValValGluThrHisAlaArgLysTyrHisValAlaLeuLeuGluHisAs 781

5046 CGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTG 5095
     ||||                                    ||||||:
 781 pGlu.....................................PheGlyI 785

5096 TAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCA 5145
     ::
 785 le.................................................. 785

5146 GGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCAC 5195

 785 ..................................................... 785

5196 ACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCA 5245

 785 ..................................................... 785

5246 AGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAG 5295
                        |||   |||:::
 786 ....................IleThrCysAsp..................... 789

5296 TGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTA 5345
        :::|||   |||
 790 ...AsnTrpArgArg................................... 793

5346 CACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCG 5395
                                    ::::::::::|||:::::::::|
 794 ..........................ValAlaValSerSerGluSerV 801
```

```
 5396 TCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGG 5445
      |||||           |||      ::::::      ||||||   |||
  801 alVal...........TyrSerAspMetAlaLysLeuArgThrLeuArg 813

 5446 GCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCT 5495
        ||||||         :::      |||::::::|||:::   |||||
  814 ArgLeuLeuLysAspGlyGluProHisValSerSerAlaLysValValLe 830

 5496 GAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACA 5545
      |
  830 u..................................................  830

 5546 AGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTG 5595

  830 .................................................. 830

 5596 AGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGG 5645

  830 .................................................. 830

 5646 CAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCG 5695
             ||||||:::|||
  831 .......ValAspGlyVal.............................. 834

 5696 ACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAA 5745

  834 .................................................. 834

 5746 GTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAA 5795

  834 .................................................. 834

 5796 GTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACG 5845
                     ||||||   ||||||
  835 ...............ProGlyCysGlyLys.................. 839

 5846 AATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGA 5895
                            |||      |||   ::::::|||:::
  840 ......................ThrLysGluIleLeuSerArgValAsn 848

 5896 AGTTCTGAGGACATAAATATGGCG.....................GTGAA 5924
        :::||||||:::   ::::::                      :::
  849 PheGluGluAspLeuIleLeuValProGlyArgGlnAlaAlaGluMetIl 865

 5925 GAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACT 5974
       :::|||      |||   :::|||:::      ||||||     |||
  865 eArgArgArgAlaAsnAlaSerGlyIleIleValAlaThrLysAspAsnV 882

 5975 CCAGGGTGGTA...AACTTTATCGTC.................AGGGGA 6003
        |||:::|||   :::|||::::::                  |||
  882 alArgThrValAspSerPheLeuMetAsnTyrGlyLysGlyAlaArgCys 898

 6004 ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTT 6053
      ::::::||||||:::::::::|||||      |||:::|||   |||
  899 GlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisThrGlyCy 915

 6054 ACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTG 6103
       :::::::::   |||      :::   ::::::   :::      :::|
  915 sValAsnPheLeuValGluMetSerLeuCysAspIleAlaTyrValTyrG 932

 6104 GAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATG 6153
      |||||   :::||||||||||:::|||||||||          |||
  932 lyAspThrGlnGlnIleProTyrIleAsnArgValThrGlyPheProTyr 948

 6154 GATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAA 6203
        :::                :::   ::::::|||
  949 ProAlaHisPheAlaLysLeuGluValAspGluValGluThrArgArgTh 965

 6204 ATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA....... 6246
       :::   |||||||||   ||||||   ::::::|||::::::
  965 rThrLeuArgCysProAlaAspValThrHisPheLeuAsnGlnArgTyrG 982

 6247 ...........ATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAG 6285
```

```
                    |||   :::   :::   :::|||         ::::::
  982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998

 6286 GTCGTTAGCGGT......AAGGACAAACCAGTAGTAAGATCGCTGTCCAA 6329
      :::|||||||||      :::   :::||||||   :::   |||:::
  999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl 1015

 6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT 6379
       :::
 1015 yLys.....................................I 1017

 6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
        |||   :::||||||         ||||||   |||:::::::|||
 1017 leLeuThrPheThrGln.........SerAsp...LysGluAlaLeuLeu 1029

 6430 GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA 6479
      ::::::|||      :::   |||   ||||||||||||:::||||||::
 1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyGl 1045

 6480 AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC 6529
      :|||::::::||||||   |||   |||          :::      |||:
 1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062

 6530 TATTCACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
      ::::::::::   :::||||||:::||||||:::|||||||||||||||
 1062 leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078

 6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
      :::||||||   |||   :::               |||   |||   ::
 1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIle:l 1095

 6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
      :   |||      :::|||:::::::|||:::   |||
 1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107
```

```
seq_name: SwissProt_39:RRPO_TOMK2

seq_documentation_block:
ID   RRPO_TOMK2     STANDARD;      PRT;  1616 AA.
AC   P89676; P90349;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain Kazakh K2) (ToMV) (TMV strain K2).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98012318; PubMed=9454068;
RA   Belenovich E.V., Generozov E.V., Novikov V.K., Zavriev S.K.;
RT   "Properties and structure of the tobacco mosaic virus strain K2
RT   genome.";
RL   Mol. Biol. (Mosk) 31:826-830(1997).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z92909; CAB07439.1; -.
DR   EMBL; Z92909; CAB07438.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
```

```
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183614 MW;  C2F5CE0C8C965336 CRC64;


alignment_scores:
            Quality:  303.00              Length:   1879
              Ratio:  0.437                Gaps:     64
 Percent Similarity:  36.935   Percent Identity: 16.711

alignment_block:
 US-09-301-906-3 x RRPO_TOMK2   ..

 Align seg 1/1  to: RRPO_TOMK2  from: 1  to: 1616

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||      :::   |||:::::::::::::::::::   ::
       2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::   :::|||   ::::::   |||              ||||||
      18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::   |||   |||:::   ||||||            ::::::|||
      35 laVal...AspGluPheAsnAlaArgAsp............ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::   :::   :::      ::::::   ::::::
      47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
          :::   :::|||||::::::|||   |||   ::::::   ::::::
      63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||   :::|||:::::::::||||||      ::::::   ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
         :::         :::         :::|||||||||||||:::|||::
      97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

    1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
         :   |||:::      |||::::::      |||   |||   |||   |
     113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

    1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.....ATC 1641
         ||||||||:::|||:::   |||:::   ::::::
     130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

    1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCC..... 1680
         |||:::||||||          |||:::::::::|||
     147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

    1681 .....AGTGACCTTCTAACTGAAGCGGCTTCAA 1708
                          :::::::::   :::|||::::::
     163 nLysGluAlaPheGluArgTyrAlaGluMetProAsnGluValValCysH 180

    1709 AGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC...GATTCTAGA 1755
           ::::::            |||:::||||||:::|||      :::|||
     180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196

    1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
         :::   |||      :::   ::::::||||||||      ::::::
     197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

    1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
           :::   |||:::   :::|||::::::   |||       |||:::
     213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

    1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
           |||      :::||||||      :::::::::|||   |||:::|||
     230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

    1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
         :::::::::      :::|||:::||||||   |||:::::::::   ::
     247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

    1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
         ::::   :::   :::   ::::::||||||:::||||||:::      :
     263 aSerGluSerThrLeuAsnHisSerHisSerTyrSerAsnIleLeuLysT 280

    2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
         ::         :::|||      :::|||:::   :::|||:::
     280 yrValCysLysThrTyrProAlaSerAsnArgGluValTyrMetLys 296

    2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
         |||:::
     297 GluPhe..................................... 298

    2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

     298 ..... 298

    2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

     298 ..... 298

    2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                          :::|||   |||:::   ::
     299 ..........................LeuValThrArgValAsnTh 305

    2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
         ::::      ::::::   |||   ||||||
     305 rTrpPheCysLysPheSerArgIleAspThrPhe..... 316

    2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                                    :::
     317 ..................................LeuPheTyr 319

    2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
         :::   |||:::|||:::   ||||||
     320 LysGlyValAlaHisLysGlyValAsp..... 328

    2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446

     328 ..... 328

    2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
              |||   :::|||   ||||||   |||   :::
     329 .....SerGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341

    2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
                :::   :::|||::::::   |||   |||:::   :::
     342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

    2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
          :::      ::::::   |||:::      |||      ::::::   :
     355 uLeuGluAspSerSerSerValAsnTyrTrpProLysMetArgAspM 372

    2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
         :::::   ::::||
     372 etValIleValPro..... 376

    2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

     377 ..........................................Le 377

    2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
```

```
     ||||||||||||   ||||||              |||::::::|||
 377 uPheAspIleSerLeuGluThr............SerLysArgThrArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
       :::|||:::          :::   :::|||   :::      |||:::
 390 ysGluValLeu.........ValSerLysAspPheValTyrThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   :::::::|||   ::::::|||   |||      |||   :::::
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu......ThrTyrSerAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::      |||:::|||
 418 nValLeuSerPheValGluSerIleArgSerArgVal............. 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
          :::::::::|||      |||           :::|||:::|||
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||:::         :::   |||
 446 SerLeuLeuGlnSerLeuSerMetThr..................... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
      ||||||
 455 .PhePhe.......................................... 456

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
                                        :::::::::   :::
 457 ................................LeuHisThrLysLeu 461

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
     ::::::|||:::   :::   :::   :::||||||   :::   |||
 462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477

3144 GAAA............GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
      |||           :::   |||:::|||:::||||||
 477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu.......... 490

3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231

 490 ................................................. 490

3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281

 490 ................................................. 490

3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331

 490 ................................................. 490

3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381

 490 ................................................. 490

3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
                        :::|||::::::|||           ::
 491 ....................AlaPheGlyAsnAlaPheProSerIleLy 500

3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
     :    :::   :::|||:::
 500 sGluArgLeuIleAsnArgLys........................... 507

3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
                                :::::   |||      :::|||
 508 ........................LeuIleLysIleThrGluAsnAla 515

3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
     |||
 516 LeuGlu.......................................... 517

3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631
```

```
 518 ..........................................I 518

3632 GCGCTGTTCCCGAG......GTTGTTTGGCGAGGGAAGAGTGTCAGCTCA 3675
       ||||||:::       |||:::::::::   :::   ||||||:::
 518 leArgValProAspLeuTyrValThrPheHisAspArgLeuValSerGlu 534

3676 TTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCT 3725
          :::   :::   :::
 535 TyrLysMetSerValAspMet.............................. 541

3726 TAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGA 3775
                                      |||   ::::::   |||:
 542 ..............................ProValLeuAspIleArgL 548

3776 ATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAA 3825
     ::       :::||||||   :::   |||
 548 ysLysMetGluGluThrGluGluMetTyr..................... 557

3826 GGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAA 3875

 557 ................................................. 557

3876 CGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGC 3925
                  :::::::::|||:::   :::|||:::   |||
 558 ............AsnAlaLeuSerGluLeuSerValLeuLysAsnSerA 570

3926 GCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACC 3975
               :::::::::   :::::::::
 570 spLysPheAspValAspValPheSerGlnMetCys.............. 581

3976 CATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCC 4025
          :::|||:::                        ||||||   ||
 582 .........GlnSerLeu.....................GluValAspPr 588

4026 GTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG 4075
     |   :::|||                                      |
 588 oMetThrAla.....................................A 592

4076 CTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTC 4125
     ||   :::::::::   |||   |||   :::::::|||:::      :::
 592 laLysValIleValAlaValMetSerAsnGluSerGlyLeuThrLeuThr 608

4126 TTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGG 4175
     |||:::|||   ::::::::::::::|||      :::
 609 PheGluGlnProThrGluAlaAsnValAlaLeuAlaLeu........... 621

4176 CAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAA 4225
                                                  |||:::
 622 .........................................GlnAspS 624

4226 GGGGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGC 4275
          :::|||:::||||||
 624 erGluLysAlaSerAspGly............................. 630

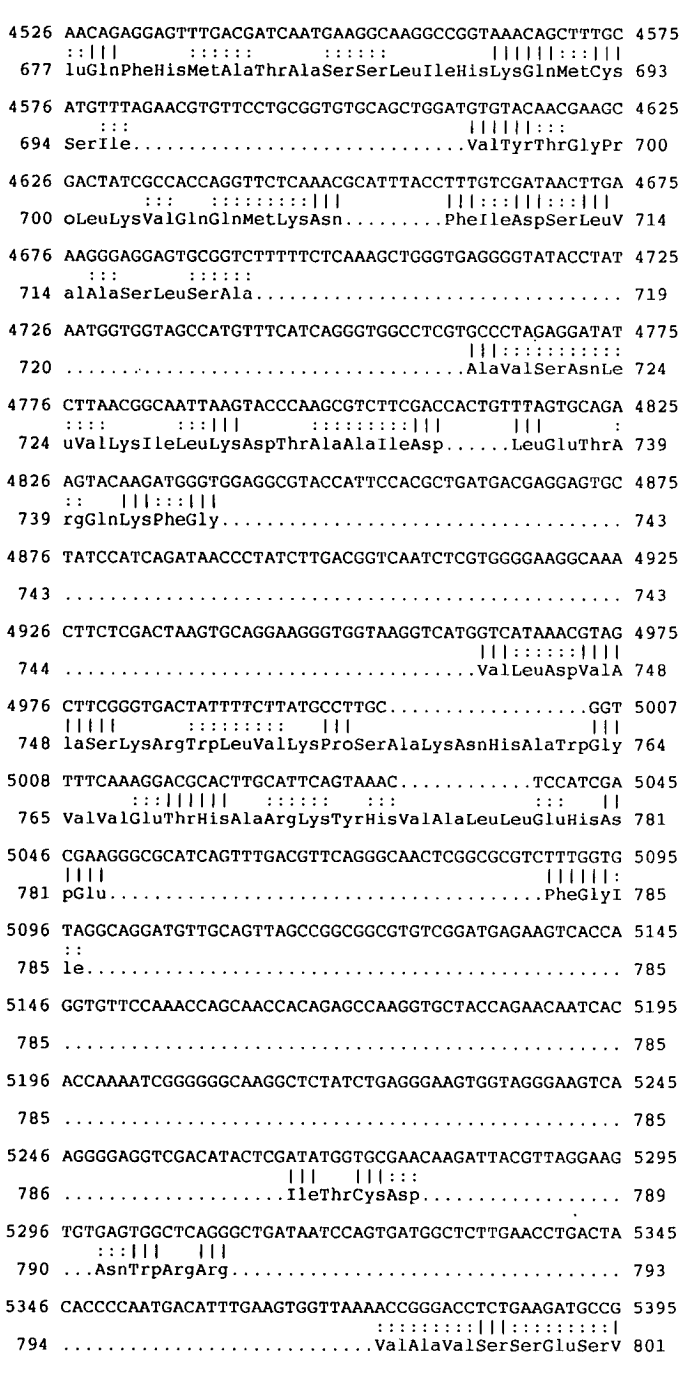
4276 GACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATAT 4325
                     :::   :::|||:::|||   :::::::::
 631 ..................AlaLeuValValThrSerArgAspValGluGl 641

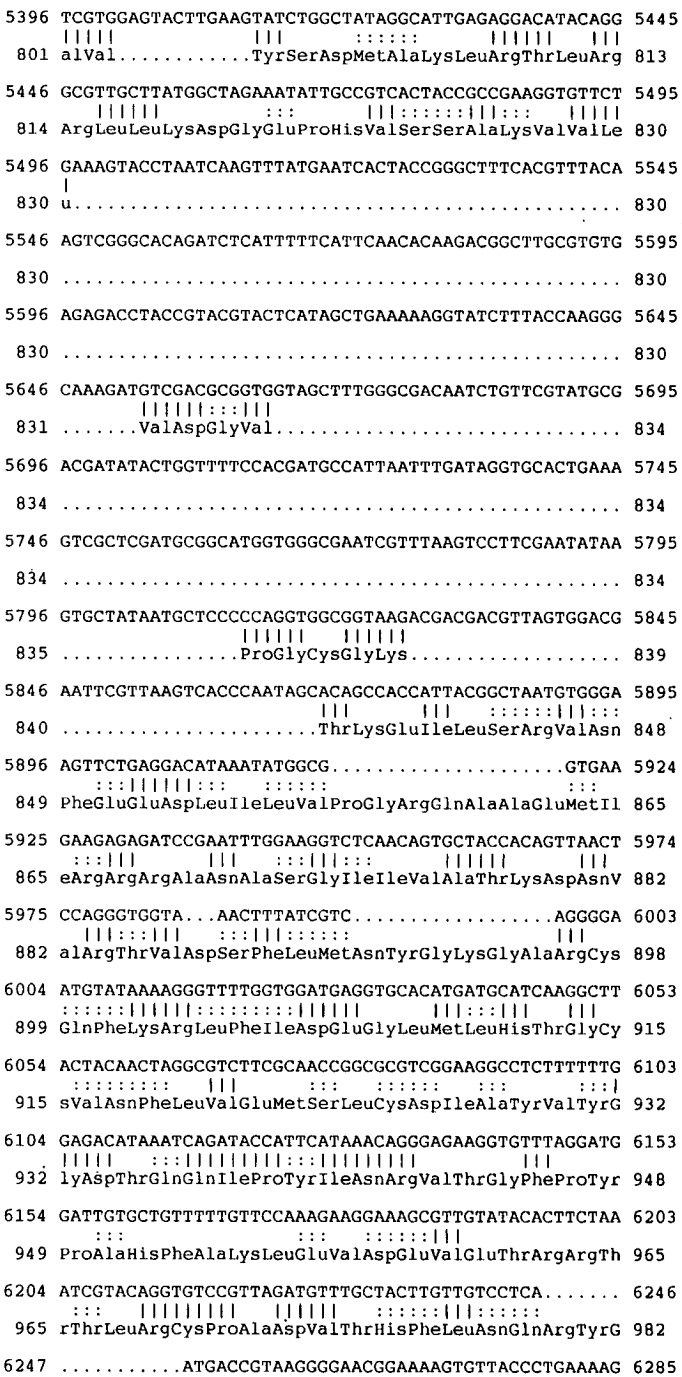
4326 TCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAG 4375
      |||::::::      |||                     ::::::   |
 641 uProSerIleArgGlySer..................MetAlaArgGlyG 652

4376 AAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425
     ||:::   ||||||   :::||||||                   :::
 652 luLeuGlnLeuAlaGlyLeuSerGly..................AspVal 662

4426 CCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAA 4475
     |||      |||   |||
 663 ProGluSerSerTyrThr.............................. 668

4476 GGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCA 4525
                              ::::::   :::::::|||:::   :
 669 .......................ArgSerGluGluIleGluSerLeuG 677
```





```
                 |||   :::   :::   :::|||          ::::::
     982 luGlyHisValMetCysThrSerSerGluLysLysSerValSerGlnGlu 998

    6286 GTCGTTAGCGGT......AAGGACAAACCAGTAGTAAGATCGCTGTCCAA 6329
         :::|||||||||      :::   :::||||||   :::   |||:::
     999 MetValSerGlyAlaAlaSerIleAsnProValSerLysProLeuLysGl 1015

    6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT 6379
           :::
    1015 yLys.........................................I 1017

    6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
           |||   :::||||||         ||||||   |||::::::|||
    1017 leLeuThrPheThrGln.........SerAsp...LysGluAlaLeuLeu 1029

    6430 GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA 6479
         :::::::|||      :::   |||   ||||||||||||:::||||||::
    1030 SerArgGly...TyrAlaAspValHisThrValHisGluValGlnGlyGl 1045

    6480 AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC 6529
         :|||::::::||||||   |||   |||         :::      |||:
    1045 uThrTyrAlaAspValSerLeuValArgLeuThrProThrProValSerI 1062

    6530 TATTCACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
         ::::::::::::   :::||||||:::||||||:::|||||||||||||||
    1062 leIleAlaArgAspSerProHisValLeuValSerLeuSerArgHisThr 1078

    6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
         :::|||||||   |||   :::                |||   |||   ::
    1079 LysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleIl 1095

    6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
         :   |||      :::|||::::::|||:::   |||
    1095 eArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107

seq_name: SwissProt_39:RRPO_TOMK1

seq_documentation_block:
ID   RRPO_TOMK1     STANDARD;      PRT;  1616 AA.
AC   Q9Q1T8; Q9Q1T7;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain Kazakh K1) (ToMV) (TMV strain K1).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196905; PubMed=10732356;
RA   Belenovich E.V., Novikov V.K., Zavriev S.K.;
RT   "Biological properties and genome structure of the Kazakh isolate K1
RT   of Tobacco Mosaic virus.";
RL   Mol. Biol. (Mosk) 34:172-176(2000).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ243571; CAB62911.1; -.
DR   EMBL; AJ243571; CAB62912.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183606 MW;  4BF9913CE791DE33 CRC64;


alignment_scores:
            Quality:  302.00              Length:   1883
              Ratio:  0.436                Gaps:     65
 Percent Similarity:  36.803   Percent Identity: 16.729

alignment_block:
 US-09-301-906-3 x RRPO_TOMK1   ..

 Align seg 1/1  to: RRPO_TOMK1  from: 1  to: 1616

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||      :::   |||::::::::::::::::::   ::
       2 AlaTyrThrGlnThrAlaThrSerSerAlaLeuLeuGluThrValArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::   :::|||   ::::::   |||                |||||
      18 yAsnAsnThrLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::   |||   |||:::   ||||||           ::::::|||
      35 laVal...AspGluPheAsnAlaArgAsp.....ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::   :::   :::      ::::::   ::::::
      47 LysValAsnPheSerLysValValSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
          :::   :::||||||::::::|||   |||   ::::::   ::::::
      63 rLysAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||   :::|||::::::::::||||||      ::::::   ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
           :::         :::         :::||||||||||||:::|||::
      97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

    1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
         :   |||:::      |||::::::      |||   |||   |||   |
     113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

    1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.....ATC 1641
         |||||||:::|||:::   |||:::   ::::::
     130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

    1642 CTCTTTCTTTCC...ACAGCTGGGGGAGATTCGTACGTGTCC..... 1680
         |||:::|||||           |||:::::::::|||
     147 LeuTyrLeuSerArgLeuGluArgGlyAsnLysHisValProAsnPheGl 163

    1681 .....AGTGACCTTCTAACTGAAGCGGCTTCAA 1708
                         :::::::::   :::|||::::::
     163 nLysGluAlaPheAspArgTyrAlaGluMetProAsnGluValValCysH 180

    1709 AGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC...GATTCTAGA 1755
           ::::::           |||:::|||||||:::|||      :::|||
     180 isAspThrPheGlnThrCysArgHisSerGlnGluCysTyrThrGlyArg 196

    1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
         :::   |||      :::   ::::::|||||||||      ::::::
     197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213
```

```
1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
 213 eGlyAlaAlaLeuLeuArgLysAsnValHisValCysTyrAlaAlaPheH 230

1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
 230 isPheSerGluAsnLeuLeuLeuGluAspSerHisValAsnLeuAspGlu 246

1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
 247 IleAsnAlaCysPheGlnArgAspGlyAspArgLeuThrPheSerPheAl 263

1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
 263 aSerGluSerThrLeuAsnTyrSerHisSerTyrSerAsnIleLeuLysT 280

2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
 280 yrValCysLysThrTyrProAlaSerAsnArgGluValTyrMetLys 296

2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
 297 GluPhe..... 298

2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152
 298 ..... 298

2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202
 298 ..... 298

2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
 299 .....LeuValThrArgValAsnTh 305

2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
 305 rTrpPheCysLysPheSerArgIleAspThrPhe..... 316

2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
 317 .....LeuLeuTyr 319

2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
 320 LysGlyValAlaHisLysGlyValAsp..... 328

2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446
 328 ..... 328

2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
 329 .....AsnGluGlnPheTyrLysAlaMetGluAspAlaTrpHis..... 341

2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
 342 .....TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
 355 uLeuGluAspSerSerSerValAsnTyrTrpPheProLysMetArgAspM 372

2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
 372 etValIleValPro..... 376

2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696
 377 .....Le 377

2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
 377 uPheAspIleSerLeuGluThr.....SerLysArgThrArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
 390 ysGluValLeu.....ValSerLysAspPheValTyrThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu.....ThrTyrSerAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
 418 nValLeuSerPheValGluSerIleArgSerArgVal..... 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
 446 SerLeuLeuGlnSerLeuSerMetThr..... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
 455 .PhePhe..... 456

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
 457 .....LeuHisThrLysLeu 461

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
 462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheAlaLeuGlyPr 477

3144 GAAA.....GAATTTTGGAGTGAAGTTTCCTTAAATGACTTCA 3181
 477 oLysThrValSerGlnHisValTrpAspGluIleSerLeu..... 490

3182 CCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCT 3231
 490 ..... 490

3232 TTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAA 3281
 490 ..... 490

3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331
 490 ..... 490

3332 CACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGT 3381
 490 ..... 490

3382 GGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAG 3431
 490 ..... 490

3432 GGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTA 3481
 490 ..... 490

3482 CCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCT 3531
 490 ..... 490

3532 CTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGC 3581
 490 ..... 490

3582 GAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGT 3631
```

```
 491 ..............................AlaPheGlyAsnAlaPheP 497

3632 GCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTA 3681
       ::::::   |||               :::   ::::::      |||
 497 roSerIleLysGlu..............ArgLeuIleAsnArgLysLeu 508

3682 CTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGA 3731
     :::   ||||||                    ::::::   |||:::
 509 IleLysIleThr.....................GluAsnAlaLeuGluIl 518

3732 CACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGT 3781
         :::|||:::|||   ||||||
 518 eArgValProAspLeu...TyrVal........................ 525

3782 ACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCT 3831

 525 ................................................. 525

3832 GGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGG 3881
               |||   |||:::   :::
 526 ............ThrPheHisAspArgLeu..................... 531

3882 ATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACG 3931
                 ||||||:::|||   :::   :::::::::      :
 532 .............ValSerGluTyrLysMetSerValAspMetProValL 544

3932 TGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCA 3981
     ::   |||      :::   |||   |||   |||
 544 euAspIleArgLysLysMetGluGluThrGluGlu.............. 555

3982 CCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGC 4031
               :::|||:::   ::::::|||
 556 ............MetTyrAsnAlaLeuSerGlu.................. 562

4032 AGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACG 4081

 562 ................................................. 562

4082 CTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCAC 4131
       ::::::   :::::::::      |||         ||||||||
 563 ..LeuSerValLeuLysThrSerAspLysPheAspValAspValPheSer 578

4132 CAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCAAAGG 4181
     |||:::               ::::::::::||||||||
 579 GlnMet..............CysGlnSerLeuGluValAsp......Pr 588

4182 GTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGGGGA. 4230
      :::         :::|||:::      :::   ::::::|||   |||
 588 oMetThrAlaAlaLysValIleValAlaValMetSerAsnGluSerGlyL 605

4231 ................AGAGCTGCAGATGGTAACACAGCACAAACCGCG 4263
                     :::   :::::::::|||:::|||      |||
 605 euThrLeuThrPheGluGlnProThrGluAlaAsnValAla...LeuAla 620

4264 GTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCC 4313
     ::::::::::::::::            :::   :::|||:::|||
 621 LeuGlnAspSerGluLysAlaSerAspGlyAlaLeuValValThrSerAr 637

4314 ACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAG 4363
      :::::::::   |||::::::      |||
 637 gAspValGluGluProSerIleArgGlySer.................M 648

4364 AAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGCCAATC 4413
       ::::::   |||:::   ||||||   :::||||||
 648 etAlaArgGlyGluLeuGlnLeuAlaGlyLeuSerGly............ 660

4414 GTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATAAT 4463
           :::   |||      |||   |||
 661 ......AspValProGluSerSerTyrThr..................... 668

4464 TGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGTACAAG 4513
                                        ::::::   ::::
 669 ..................................ArgSerGluGluI 673
```

```
4514 TCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGGCCGGT 4563
     ::|||:::   :::|||      ::::::      ::::::
 673 leGluSerLeuGluGlnPheHisMetAlaThrAlaSerSerLeuIleHis 689

4564 AAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGT 4613
     ||||||:::|||   :::                              ||
 690 LysGlnMetCysSerIle..............................Va 696

4614 GTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTG 4663
     ||||:::            :::   :::::::::|||         |||:
 696 lTyrThrGlyProLeuLysValGlnGlnMetLysAsn.........PheI 710

4664 TCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAG 4713
     ::|||:::|||   :::      ::::::
 710 leAspSerLeuValAlaSerLeuSerAla..................... 719

4714 GGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGC 4763
                                                      ||
 720 ..................................................Al 720

4764 CCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACT 4813
     |:::::::::::::::      :::|||      :::::::::|||
 720 aValSerAsnLeuValLysIleLeuLysAspThrAlaAlaIleAsp.... 735

4814 GTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGAT 4863
       |||      :::   |||:::|||
 736 ..LeuGluThrArgGlnLysPheGly........................ 743

4864 GACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGT 4913

 743 ................................................. 743

4914 GGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGG 4963
                                                     |
 744 .................................................V 744

4964 TCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGC......... 5004
     ||::::::|||||||||      :::::::::   |||
 744 alLeuAspValAlaSerLysArgTrpLeuValLysProSerAlaLysAsn 760

5005 .........GGTTTTCAAAGGACGCACTTGCATTCAGTAAAC........ 5037
             |||      :::||||||   ::::::   :::
 761 HisAlaTrpGlyValValGluThrHisAlaArgLysTyrHisValAlaLe 777

5038 ....TCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGC 5083
         :::   ||||||
 777 uLeuGluHisAspGlu................................ 782

5084 GCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGAT 5133
       ||||||:::
 783 .....PheGlyIle.................................. 785

5134 GAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTAC 5183

 785 ................................................. 785

5184 CAGAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTG 5233

 785 ................................................. 785

5234 GTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGAT 5283
                                     |||   |||:::
 786 ..............................IleThrCysAsp...... 789

5284 TACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCT 5333
                  :::|||   |||
 790 ..............AsnTrpArgArg...................... 793

5334 TGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCT 5383
                                                :::::::::|
 794 ....................................ValAlaValS 797
```

```
5384 CTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAG 5433
     ||:::::::::|||||            |||      ::::::
 797 erSerGluSerValVal.....TyrSerAspMetAlaLysLeu 809

5434 AGGACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGC 5483
     ||||||   |||   ||||||        :::      |||:::::||
 810 ArgThrLeuArgArgLeuLeuLysAspGlyGluProHisValSerSerAl 826

5484 CGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCT 5533
     |:::   ||||||
 826 aLysValValLeu..................................... 830

5534 TTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGAC 5583

 830 ..................................................... 830

5584 GGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTAT 5633

 830 ..................................................... 830

5634 CTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATC 5683
                    ||||||:::|||
 831 .................ValAspGlyVal................... 834

5684 TGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATA 5733

 834 ..................................................... 834

5734 GGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTC 5783

 834 ..................................................... 834

5784 CTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGA 5833
                                 ||||||   ||||||
 835 ..........................ProGlyCysGlyLys....... 839

5834 CGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACG 5883
                                          |||      |||
 840 ...............................ThrLysGluIleLeu 844

5884 GCTAATGTGGGAAGTTCTGAGGACATAAATATGGCG.............. 5919
     ::::::|||:::   :::||||||:::   ::::::
 845 SerArgValAsnPheGluGluAspLeuIleLeuValProGlyArgGlnAl 861

5920 .......GTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTA 5962
            :::   :::|||      |||   :::|||:::      ||||
 861 aAlaGluMetIleArgArgArgAlaAsnAlaSerGlyIleIleValAlaT 878

5963 CCACAGTTAACTCCAGGGTGGTA...AACTTTATCGTC............ 5997
     ||      |||   |||:::|||   :::|||::::::
 878 hrLysAspAsnValArgThrValAspSerPheLeuMetAsnTyrGlyLys 894

5998 ......AGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGAT 6041
           |||   ::::::||||||:::::::::||||||      |||::
 895 GlyAlaArgCysGlnPheLysArgLeuPheIleAspGluGlyLeuMetLe 911

6042 GCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAG 6091
     :|||   |||   :::::::::   |||      :::   ::::::   :
 911 uHisThrGlyCysValAsnPheLeuValGluMetSerLeuCysAspIleA 928

6092 GCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAG 6141
     ::      :::||||||   :::||||||||||:::||||||||||
 928 laTyrValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValThr 944

6142 GTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGT 6191
        |||          :::                 :::   ::::::|||
 945 GlyPheProTyrProAlaHisPheAlaLysLeuGluValAspGluValGl 961

6192 ATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGT 6241
                  :::   |||||||||   ||||||   ::::::|||:
 961 uThrArgArgThrThrLeuArgCysProAlaAspValThrHisPheLeuA 978

6242 CCTCA.................ATGACCGTAAGGGGAACGGAAAAGTGT 6273
```

```
     :::::                    |||   :::   :::   :::|||
 978 snGlnArgTyrGluGlyHisValMetCysThrSerSerGluLysLysSer 994

6274 TACCCTGAAAAGGTCGTTAGCGGT......AAGGACAAACCAGTAGTAAG 6317
          :::::::::|||||||||.|      :::   :::||||||  ::
 995 ValSerGlnGluMetValSerGlyAlaAlaSerIleAsnProValSerLy 1011

6318 ATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAA 6367
     :   |||:::   :::
1011 sProLeuLysGlyLys................................. 1016

6368 ACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAG 6417
                  |||   :::||||||         ||||||   |||
1017 ..........IleLeuThrPheThrGln.........SerAsp...Lys 1025

6418 AGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGA 6467
     :::::::|||   :::::::|||      :::   |||   ||||||||||
1026 GluAlaLeuLeuSerArgGly...TyrAlaAspValHisThrValHisGl 1041

6468 AGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAG 6517
     |:::||||||:::|||:::::::||||||   |||   |||        :
1041 uValGlnGlyGluThrTyrAlaAspValSerLeuValArgLeuThrProT 1058

6518 CCGATGACTCCCTATTCACTAAA...CAACCGCATATACTTGTTGGTTTG 6564
     ::      |||:::::::::::   :::||||||:::|||||||:::|||
1058 hrProValSerIleIleAlaArgAspSerProHisValLeuValSerLeu 1074

6565 TCGAGACACACACGCTCACTGGTTTATGCCGCT..............CT 6599
     ||||||||||||:::||||||   |||   :::               ||
1075 SerArgHisThrLysSerLeuLysTyrTyrThrValValMetAspProLe 1091

6600 GAGCTCAAAGTTGGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
     |   |||   :::   |||      :::|||::::::|||:::   |||
1091 uValSerIleIleArgAspLeuGluArgValSerSerTyrLeuLeuAsp 1107
```

```
seq_name: SwissProt_39:RRPO_CGMVS

seq_documentation_block:
ID   RRPO_CGMVS     STANDARD;      PRT;  1648 AA.
AC   P19523; Q83208; P89877; P90356;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
OS   Cucumber green mottle mosaic virus (watermelon strain SH) (CGMMV), and
OS   Cucumber green mottle mosaic virus (watermelon strain W) (CGMMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SH;
RX   MEDLINE=91311400; PubMed=1856687;
RA   Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA   Sato T., Motoyoshi F., Nishiguchi M.;
RT   "The complete nucleotide sequence of cucumber green mottle mosaic
RT   virus (SH strain) genomic RNA.";
RL   J. Gen. Virol. 72:1487-1495(1991).
RN   [2]
RP   REVISIONS.
RC   STRAIN=SH;
RA   Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA   Sato T., Motoyoshi F., Nishiguchi M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1496-1648 FROM N.A.
RC   STRAIN=W;
RX   MEDLINE=89073773; PubMed=3201760;
RA   Saito T., Imai Y., Meshi T., Okada Y.;
RT   "Interviral homologies of the 30K proteins of tobamoviruses.";
RL   Virology 167:653-656(1988).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
```

```
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN
CC       CODONS FOR LYS-1144 AND GLN-1146.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D12505; BAA18895.1; -.
DR   EMBL; D12505; BAA18896.1; -.
DR   EMBL; J04322; AAA46382.1; -.
DR   PIR; JQ1157; WMTMS2.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1648       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1144       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     863    870       ATP (POTENTIAL).
SQ   SEQUENCE   1648 AA;  186549 MW;  1D1AFFEEE7B65595 CRC64;


alignment_scores:
            Quality:  298.00              Length:   1919
              Ratio:  0.434                Gaps:     72
 Percent Similarity:  35.800   Percent Identity: 17.405

alignment_block:
 US-09-301-906-3 x RRPO_CGMVS   ..

 Align seg 1/1  to: RRPO_CGMVS  from: 1  to: 1648

    1213 AATCAAGAGGAGGTCTATTCATCG.....GTCGTTTCGAGTAT 1250
         ||||||:::::::::   :::|||          :::||||||:::::
      10 AsnGlnArgAspAlaAlaAlaSerGlyArgAsnAsnLeuValSerGlnLe 26

    1251 CACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGG 1300
         ::::::::::   |||      ::            |||:::|||   |
      26 uAlaSerLysArgValTyrAspGluAlaValArgSerLeuAspHisGlnA 43

    1301 ACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAA 1350
         ||                                          :::
      43 sp.........................................Arg 44

    1351 AAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAA 1400
         :::|||:::   :::   :::      ::::::   ::::::   :::
      45 ArgProLysMetAsnPheSerArgValValSerThrGluHisThrArgLe 61

    1401 GCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCA 1450
          :::   :::   :::||||||:::||||||:::|||:::   :::   :
      61 uValThrAspAlaTyrProGluPheSerIleSerPheThrAlaThrLysA 78

    1451 GGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATC 1500
         ::|||   |||   :::|||::::::::::|||
      78 snSerValHisSerLeuAlaGlyGlyLeuArgLeuGluLeuGluTyr 94

    1501 TTTTCCAGGAGGTGT......GGTAATGTGTGCTTCTTCGATATTGGGGG 1544
         :::      :::        |||:::   |||   :::||||||||||
      95 MetMetMetGlnValProTyrGlySerProCys...TyrAspIleGlyGl 110

    1545 GAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCA 1594
         |:::::|||   |||:::      |||:::         |||   ||||
     110 yAsnTyrThrGlnHisLeuPheLysGlyArgSerTyrValHisCysCysA 127

    1595 ATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATC......AATGAG 1638
         |||||   ||||||:::|||||||||   |||::::::      |||:::

    1639 ATCCTCTTT..................................... 1647
         ::::::
     144 MetIleThrGlnHisValGlnArgHisLysGlySerGlyGlyCysArgPr 160

    1648 .CTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTG 1696
          |||   |||         |||::::::         |||
     160 oLeuProThrPheGlnIleAspAlaPheArgArgTyrAsp......... 173

    1697 AAGCGGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGC 1746
              :::|||      :::      ||||||         |||:::|||
     174 .....SerSerProCysAlaValThrCysSerAspValPheGlnGluCys 188

    1747 GATTCTAGAGCCGATGCGGGT.................TTTATGGTGGA 1778
         :::          :::|||                        :::   ::
     189 SerTyrAspPheGlySerGlyArgAspAsnHisAlaValSerLeuHisSe 205

    1779 TGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGG 1828
         ::::|||||||||      ::::::::::   |||:::   :::|||:
     205 rIleTyrAspIleProTyrSerSerIleGlyProAlaLeuHisArgLysA 222

    1829 GTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTAC 1878
         :::::   |||         |||:::   |||         ||||||
     222 snValArgValCysTyrAlaAlaPheHisPheSerGluAlaLeuLeuLeu 238

    1879 GGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGA 1928
         |||:::          |||::::::::::   :::      :::   ::
     239 GlySerProValGlyAsnLeuAsnSerIleGlyAlaGlnPheArgValAs 255

    1929 AGGTGATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATG...TATG 1975
         :||||||   :::   :::      ::::::   :::   :::   |||
     255 pGlyAspAspValHisPheLeuPheSerGluGluSerThrLeuHisTyrT 272

    1976 AACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGC 2025
           ||||||::::::|||::::::   :::         :::|||
     272 hrHisSerLeuGluAsnIleLysLeuIleValMetArgThrTyrPhePro 288

    2026 ACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGG 2075
         :::::::::      |||:::      |||:::         |||
     289 AlaAspAspArgPheValTyrIleLysGluPheMetValLysArgValAs 305

    2076 TTACCATCATCTCACTATGTGTAGGGCT..............CAGAAGT 2110
                  :::   :::   ||||||              :::||||
     305 pThrPhePhePheArgLeuValArgAlaAspThrHisMetLeuHisLysS 322

    2111 CACCTGGA..................ACTGAGGTTACGTATAGGTCGTTG 2142
         ||   |||                   :::|||            :::
     322 erValGlyHisTyrSerLysSerLysSerGluTyrPheAlaLeuAsnThr 338

    2143 GTCCCGTCGTTCGTGGGCAAATCGCTG..................... 2169
            |||   |||      |||:::
     339 ProProIlePheGlnAspLysAlaThrPheSerValTrpPheProGluAl 355

    2170 ..........GTGTTCATACCTGTT..............GTAGCTGGTT 2194
                   |||:::||||||                    ::::
     355 aLysArgLysValLeuIleProLysPheGluLeuSerArgPheLeuSerG 372

    2195 CTAGTGTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGG 2244
         :::::|||:::::::::   :::::::::|||:::||||||||
     372 lyAsnValLysIleSerArgMetLeuValAspAlaAspPheValHisThr 388

    2245 ATCTATTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGA..... 2289
         |||            :::|||   |||:::|||::::::|||:::
     389 Ile............IleAsnHisIleSerThrTyrAspAsnLysAlaLe 401

    2290 .ACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCA 2338
          :::::::::   :::   :::   |||:::|||   ::::::::||||
     401 uValTrpLysAsnValGlnSerPheValGluSerIleArgSerArgValI 418

    2339 TTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGAT 2388
         ||:::::::::      |||   |||:::   ::::::      |||:::
     418 leValAsnGlyValSerValLysSerGluTrpAsnValProValAspGln 434
```

```
2389 ATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGC 2438
     :::      :::      :::::::::        :::|||   |||
 435 LeuThrAspIleSerPheSerIlePheLeu...LeuValLysValArgLy 450

2439 GAAGAGTATTCGC......TCCTATAACTTTATAAAAGCCAGTGAGGGGA 2482
       :::|||:::      |||   :::   :::       ::::::|||
 450 sValGlnIleGluLeuMetSerAspLysValValIleGluAlaArgGlyL 467

2483 GTCTC.........GCCGGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGC 2523
       |||          |||       :::|||          :::   :::|||
 467 euLeuArgArgPheAlaAspSerLeuLysSerAlaValGluGlyLeuGly 483

2524 GATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGA 2573
     ||||||                  ||||||
 484 AspCys.....ValTyr..... 487

2574 TAACTTCAACGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCATCC 2623

 487 ..... 487

2624 GTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTTCA 2673

 487 ..... 487

2674 GACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTCGG 2723
     |||   |||        |||   ||||||   |||
 488 AspAlaLeuValGlnThrGlyTrpPheAspThrSer..... 499

2724 TAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGG 2773
                                             :
 500 .....S 500

2774 AAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCC 2823
     ::|||::::::||||||:::                    |||
 500 erAspGluLeuLysValLeu.....LeuPro 508

2824 CTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTC 2873
       |||:::       |||   |||
 509 GluProPheMetThrPheSerAspTyrLeu..... 518

2874 TAACGTCCACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCG 2923
                        |||:::      ::::::
 519 .................GluGlyMetTyrGluAlaAspAlaLysIleG 529

2924 TGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAG 2973
       :::||||||||||:::   |||::::::                |||
 529 luArgGluSerValSerGluLeuLeuAlaSerGlyAspAspLeuPheLys 545

2974 GCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGA 3023
       ::::::   |||   :::   :::|||
 546 LysIleAspGluIleArgAsnAsnTyrSer..... 555

3024 CCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGA 3073
       |||:::      |||
 556 ....GlyValGluPheAsp..... 560

3074 ATGCTATAAACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTG 3123

 560 ..... 560

3124 AGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAA 3173
             |||         :::||||||   :::|||:::::::::
 561 .........ValGluLysPheGlnGluPheCysLysGluLeuAsnVal.. 573

3174 TGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCG 3223
                              |||   :::||||||||||
 574 .......................AsnProMetLeuIleGly..... 579

3224 TGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATT 3273

 579 ..... 579
```

```
3274 GCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTC 3323
                                          ::::::::::::::
 580 ..................................HisValIleGluAl 584

3324 GTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAG 3373
       ::::   ||||||   ||||||
 584 aIlePheSerGlnLysAlaGly..... 591

3374 TATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACC 3423
                                                  ||||||
 592 .....ValThr 593

3424 GTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATC 3473
     ||||||
 594 ValThr..... 595

3474 GTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGG 3523

 595 ..... 595

3524 CACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTG 3573
                                        ||||||||||||||||
 596 .....GlyLeuGlyThrLeu 600

3574 CTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAA 3623

 600 ..... 600

3624 TAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCT 3673
                    ||||||:::         |||   ||||||:::
 601 .............SerProGluMet.........GlyAlaSerValAla. 609

3674 CATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTG 3723

 609 ..... 609

3724 CTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAG 3773
     |||:::::::|||
 610 LeuSerSerThr..... 613

3774 GAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTATGACG 3823

 613 ..... 613

3824 AAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGAT 3873

 613 ..... 613

3874 AACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGT 3923
            |||::::::::::   ::::::   ||||||:::   :::::
 614 .........SerValAspThrCysGluAspMetAspValThrGluAspMe 627

3924 GCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGA 3973
     ::::|||:::                     |||         :::::::
 627 tGluAspIle.....ValLeuMetAlaAspLysS 637

3974 CCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCC 4023
     ::||||||                  |||::::::   |||:::|||
 637 erHisSer.....TyrMetSerProGluMetAla 646

4024 CCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTAC 4073
          |||
 647 ArgTrpAla..... 649

4074 CGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATG 4123
                              :::||||||   |||
 650 .....AspValLysTyrGly..... 654

4124 TCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGAC 4173
                                                    :::
 655 .....Asn 655

4174 GGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACA..... 4218
```

```
     :::||||||   |||   |||      ||||||   :::   |||
 656 AsnLysGlyAlaLeuValGluTyrLysValGlyThrSerMetThrLeuPr 672

4219 .........CAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACAAA 4258
                  |||:::|||:::|||:::
 672 oAlaThrTrpAlaGluLysGlyLysAlaVal..................... 682

4259 CCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGAGT 4308
                               |||:::::::::      |||
 683 .....................LeuProLeuSerGlyIleCysValArg 691

4309 TCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCTCA 4358
     :::||||||                                      ::
 692 LysProGln.................................PheSe 696

4359 GAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCACGC 4408
     :|||   :::   :::|||
 696 rLysProLeuAspGluGlu.............................. 702

4409 CAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAG 4458
            |||:::            :::::::::         |||:::
 703 ........AspAspLeuArgLeuSerAsnMetAsnPhePheLysValSer 716

4459 ATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAGGT 4508
        :::          |||::::::::::   |||:::           ::
 717 AspLeuLysLeuLysLysThrIleThrProValValTyrThrGlyThrIl 733

4509 ACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAAGG 4558
     ::::   :::|||   |||
 733 eArgGluArgGlnMetLys.............................. 739

4559 CCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTG 4608

 739 ..................................................... 739

4609 GATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTAC 4658
                                                      ::
 740 ..................................................As 740

4659 CTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGG 4708
     :::::::|||   |||         ||||||   :::   |||   ||||
 740 nTyrIleAspTyrLeu.........SerAlaSerLeuGlySerThrLeuG 754

4709 GTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCT 4758
     ||               |||             |||   |||   |||
 754 ly..............AsnLeuGluArgIleValArgSerAspTrpAsn 765

4759 CGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGA 4808
        :::   |||::::::   |||         |||
 766 GlyThrGluGluSerMetGlnThrPheGlyLeuTyr.............. 777

4809 CCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACG 4858

 777 ..................................................... 777

4859 CTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAAT 4908
       |||   |||:::|||
 778 ..AspCysGluLysCys................................. 782

4909 CTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGT 4958

 782 ..................................................... 782

4959 CATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTT 5008
                                  ::::::|||:::|||:::
 783 ........................LysTrpLeuLeuLeuProAla.... 789

5009 TTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATC 5058
       ::::::   |||      :::|||                    :::
 790 ..GluLysLysHisAlaTrpAlaVal.....................Val 798

5059 AGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
           :::      :::||||||   :::|||
```

```
 799 LeuAlaSerAspAspThrThrArgIleIlePhe.................. 809

5109 GCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACC 5158
                     :::|||   ::::::|||   |||   |||
 810 ...............LeuSerTyrAspGluSer...GlySerPro.... 818

5159 AGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGG 5208
                                        |||      |||::::::
 819 ..............................IleIleAspLysLysAsn 824

5209 GGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGAC 5258
        |||   ::::::      |||      :::|||
 825 TrpLysArgPheAlaValCysSerGluThrLysVal.............. 836

5259 ATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCA 5308
       ||||||:::                    :::|||:::   |||   |||
 837 .TyrSerVal.................IleArgSerLeuGluValLeu. 846

5309 GGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACA 5358

 846 ..................................................... 846

5359 TTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTT 5408
                               :::::::::|||:::|||:::
 847 ......................AsnLysGluAlaIleValAsp..... 853

5409 GAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGG 5458

 853 ..................................................... 853

5459 CTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT 5508

 853 ..................................................... 853

5509 CAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGA 5558
                       ||||||   |||:::
 854 .................ProGlyValHisIle................. 858

5559 TCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGT 5608
                  :::      ||||||:::
 859 ............ThrLeuValAspGlyVal.................... 864

5609 ACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGAC 5658

 864 ..................................................... 864

5659 GCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGT 5708

 864 ..................................................... 864

5709 TTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCG 5758

 864 ..................................................... 864

5759 GCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCT 5808

 864 ..................................................... 864

5809 CCCCCAGGTGGCGGTAAGACGACGACGTTAGTG.................. 5841
       ||||||   |||||||||:::   ::::::
 865 ...ProGlyCysGlyLysThrAlaGluIleIleAlaArgValAsnTrpLy 880

5842 .GACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAATG 5890
        :::::::|||   :::|||:::      |||:::
 880 sThrAspLeuValLeuThrProGlyArgGluAlaAla.............. 892

5891 TGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAAT 5940
                                     :::::::::|||
 893 ......................AlaMetIleArgArgArgAlaCysAla 901

5941 TTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTT 5990
     |||:::            ||||||:::   |||   |||:::   :::
 902 LeuHisLysSerProValAlaThrAsnAspAsnValArgThrPheAspSe 918
```

```
 5991 TATCGTC......AGGGGAATGTATAAA......AGGGTTTTGGTGGATG 6028
       :::|||      |||   :::::::|||         |||   |||||||
  918 rPheValMetAsnArgLysIlePheLysPheAspAlaValTyrValAspG 935

 6029 AGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACC 6078
      ||      |||:::|||   |||||||||:::::::::::      :::
  935 luGlyLeuMetValHisThrGlyLeuLeuAsnPheAlaLeuLysIleSer 951

 6079 GGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCAT 6128
      |||:::::::::::::::   |||||||||   :::|||||||||||||||
  952 GlyCysLysLysAlaPheValPheGlyAspAlaLysGlnIleProPheIl 968

 6129 AAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGA 6178
      |||||||      |||::::::::::|||              ||||||:
  968 eAsnArg......ValMetAsnPheAsp............TyrProLysG 979

 6179 AG.................GAAAGCGTTGTATACACTTCTAAATCGTAC 6210
      ::                 ::::::|||               ::::::
  979 luLeuArgThrLeuIleValAspAsnValGluArgArgTyrValThrHis 995

 6211 AGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGG 6260
      |||||||||   ||||||      :::|||::::::::::
  996 ArgCysProArgAspValThrSerPheLeuAsnThrIle........... 1008

 6261 AACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAG 6310
                          |||::::::   :::      :::||||
 1009 ..................TyrLysAlaAlaValAlaThrThrSerProV 1019

 6311 TAGTAAGATCGCTGTCCAAA.........................AGG 6333
      |||||:::|||::::::                            |||
 1019 alValHisSerValLysAlaIleLysValSerGlyAlaGlyIleLeuArg 1035

 6334 CCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTT 6383
      |||                 ::::::::::::|||:::::::   :::   ::
 1036 Pro..............GluLeuThrLysIleLysGlyLysIle...Il 1046

 6384 GTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAA 6433
      :   :::||||||   :::|||:::::::::   :::|||
 1046 eThrPheThrGlnSerAspLysGlnSerLeuIleLysSer.......... 1059

 6434 AAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACA 6483
                :::   |||   ||||||||||||   ||||||:::|||
 1060 .....GlyTyrAsnAspValAsnThrValHisGluIleGlnGlyGluThr 1074

 6484 TTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATT 6533
      |||:::::::::::::::   |||:::      :::      :::||||::
 1075 PheGluGluThrAlaValValArgAlaThrProThrProIleGlyLeuIl 1091

 6534 CACTAAA...CAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCT 6580
      :::::::   :::||||||:::||||||:::|||:::|||||||||:::::
 1091 eAlaArgAspSerProHisValLeuValAlaLeuThrArgHisThrLysA 1108

 6581 CACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACA 6630
      :::::||||||   ::::::         :::   |||   |||   :::
 1108 laMetValTyrTyrThrVal.........ValPheAspAlaValThrSer 1121

 6631 TATATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACAC 6680
         |||:::|||:::      ::::::|||:::::::::::::|||
 1122 IleIleAlaAspVal......GluLysValAspGlnSerIleLeuThrMe 1136

 6681 GTTCGCC 6687
       ||||||
 1136 tPheAla 1138

seq_name: SwissProt_39:RRPO_PMMVJ

seq_documentation_block:
ID   RRPO_PMMVJ     STANDARD;      PRT;  1612 AA.
AC   P89657; P90347;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Pepper mild mottle virus (strain Japan) (PMMV-J).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kirita M., Akutsu K., Watanabe Y., Tsuda S.;
RT   "Nucleotide sequence of the Japanese isolate of pepper mild mottle
RT   tobamovirus.";
RL   Ann. Phytopathol. Soc. Jpn. 0:0-0(1997).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB000709; BAA19167.1; -.
DR   EMBL; AB000709; BAA19166.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1612       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1117       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     835    842       ATP (POTENTIAL).
SQ   SEQUENCE   1612 AA;  183303 MW;  E7E66CE8ED81FE4C CRC64;


alignment_scores:
            Quality:  297.00              Length:    854
              Ratio:  0.754                 Gaps:     40
 Percent Similarity:  46.136   Percent Identity: 22.834

alignment_block:
 US-09-301-906-3 x RRPO_PMMVJ   ..

 Align seg 1/1  to: RRPO_PMMVJ  from: 1  to: 1612

 4522 CCCAAACAGAGGAGTTTGACGATC..........................  4545
      ||||||:::::::::::::::|||
  366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuGl  382

 4546 .AATGAAGGCAAG......GCCGGTAAACAGCTTTGCATGTTTAGAACGT  4588
       ||||||||||||      |||   |||::::::   :::   :::
  382 nAsnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAspP  399

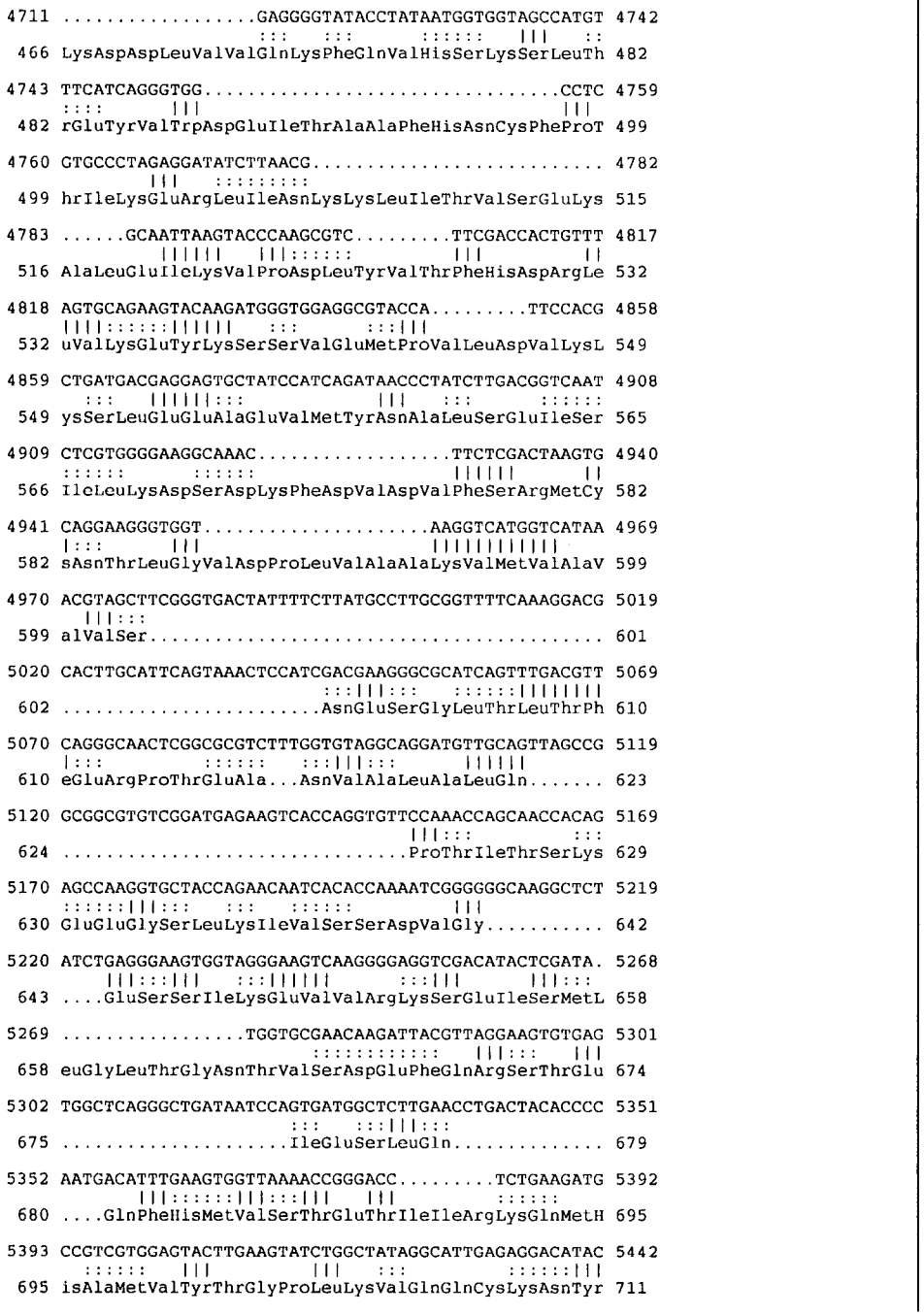
 4589 GTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACC  4638
                 |||         :::   :::      :::   |||
  399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu  415

 4639 AGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGG.....  4683
         ::::::|||::::::::::||||||:::::::::::::::|||
  416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValIl  432

 4683 ..................................................  4683

  432 eIleAsnGlyValThrAlaArgSerGluTrpAspValAspLysAlaLeuL  449

 4684 ..........AGTGCGGTCTTTTTC......TCAAAGCTGGGT......  4710
                 |||   :::||||||      :::||||||:::
  449 euGlnSerLeuSerMetThrPhePheLeuGlnThrLysLeuAlaMetLeu  465
```


4711 GAGGGGTATACCTATAATGGTGGTAGCCATGT 4742
466 LysAspAspLeuValValGlnLysPheGlnValHisSerLysSerLeuTh 482
4743 TTCATCAGGGTGG.....CCTC 4759
482 rGluTyrValTrpAspGluIleThrAlaAlaPheHisAsnCysPheProT 499
4760 GTGCCCTAGAGGATATCTTAACG..... 4782
499 hrIleLysGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys 515
4783 GCAATTAAGTACCCAAGCGTC.....TTCGACCACTGTTT 4817
516 AlaLeuGluIleLysValProAspLeuTyrValThrPheHisAspArgLe 532
4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCA.....TTCCACG 4858
532 uValLysGluTyrLysSerServalGluMetProValLeuAspValLysL 549
4859 CTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAAT 4908
549 ysSerLeuGluGluAlaGluValMetTyrAsnAlaLeuSerGluIleSer 565
4909 CTCGTGGGGAAGGCAAAC.....TTCTCGACTAAGTG 4940
566 IleLeuLysAspSerAspLysPheAspValAspValPheSerArgMetCy 582
4941 CAGGAAGGGTGGT.....AAGGTCATGGTCATAA 4969
582 sAsnThrLeuGlyValAspProLeuValAlaAlaLysValMetValAlaV 599
4970 ACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACG 5019
599 alValSer..... 601
5020 CACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTT 5069
602 AsnGluSerGlyLeuThrLeuThrPh 610
5070 CAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
610 eGluArgProThrGluAla...AsnValAlaLeuAlaLeuGln..... 623
5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAG 5169
624 ProThrIleThrSerLys 629
5170 AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCT 5219
630 GluGluGlySerLeuLysIleValSerSerAspValGly..... 642
5220 ATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATA. 5268
643 GluSerSerIleLysGluValValArgLysSerGluIleSerMetL 658
5269 TGGTGCGAACAAGATTACGTTAGGAAGTGTGAG 5301
658 euGlyLeuThrGlyAsnThrValSerAspGluPheGlnArgSerThrGlu 674
5302 TGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCC 5351
675 IleGluSerLeuGln..... 679
5352 AATGACATTTGAAGTGGTTAAAACCGGGACC.....TCTGAAGATG 5392
680 GlnPheHisMetValSerThrGluThrIleIleArgLysGlnMetH 695
5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC 5442
695 isAlaMetValTyrThrGlyProLeuLysValGlnGlnCysLysAsnTyr 711
5443 ...AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGG 5489
712 LeuAspSerLeuValAlaSerLeuSerAlaAlaValSerAsnLeuLysLy 728
5490 TGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACG 5539
728 sIleIleLys..... 731
5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTG 5589
732 AspThrAlaAlaIleAspLeuGluThrLysGluLysPhe 744
5590 CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATC..... 5634
745 GlyVal.....TyrAspValCysLeuLysLysTrpLeuValLy 757
5635 TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACA 5680
757 sProLeuSerLysGlyHisAlaTrpGlyValValMet.....AspS 771
5681 ATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTG 5730
771 erAspTyrLysCysPheValAlaLeuLeuThrTyrAspGlyGluAsnIle 787
5731 ATAGGTGCACTGAAAGTCGCTCGATGCGGC.....ATGGTGGG 5768
788 Val.....CysGlyGluThrTrpArgArgValAl 797
5769 CGAATCGTTTAAGTCCTTCGAATATAAGTGCTAT..... 5802
797 aValSerSerGluSerLeuValTyrSerAspMetGlyLysIleArgAlaI 814
5803 AATGCTCCCCCAGGTGGCGGTAAGACGACG 5832
814 leArgSerValLeuLysAspGlyGluProHisIleSerSerAlaLysVal 830
5833 ACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTAC 5882
831 ThrLeuValAspGly...ValProGlyCysGlyLysThrLysGluIleIle 846
5883 GGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAG.... 5928
846 uSerArgValAsnPheAspGluAspLeuValLeuValProGlyLysGlnA 863
5929 AGAGATCCGAATTTGGAAGGTCTCAACAGTGCT 5961
863 laAlaGluMetIleArgArgArgAlaAsnSerSerGlyLeuIleValAla 879
5962 ACCACAGTTAACTCCAGGGTGGTA...AACTTTATCGTC.....AG 5999
880 ThrLysGluAsnValArgThrValAspSerPheLeuMetAsnTyrGlyAr 896
6000 GGGA.....ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGC 6043
896 gGlyProCysGlnTyrLysArgLeuPheLeuAspGluGlyLeuMetLeuH 913
6044 ATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGC 6093
913 isProGlyCysValAsnPheLeuValGlyMetSerLeuCysSerGluAla 929
6094 CTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGT 6143
930 PheValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValAlaTh 946
6144 GTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTAT 6193
946 rPheProTyrProLysHisLeuSerGlnLeuGluValAspAlaValGluT 963
6194 ACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
963 hrArgArgThrThrLeuArgCysProAlaAspIleThrPhePheLeuAsn 979
6244 TCA.....ATGACCGTAAGGGGAACGGAAAAGTGTTA 6275

```
      :::                  |||   :::   ::::::   :::
  980 GlnLysTyrGluGlyGlnValMetCysThrSerSerValThrArgSerVa 996

 6276 CCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGT 6325
         ::::::|||::::::|||            |||::::::   :::|
  996 lSerHisGluValIleGlnGly......AlaAlaValMetAsnProValS 1011

 6326 CCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGAC 6375
      |||||         |||   :::
 1011 erLysProLeuLysGlyLysVal..........................  1018

 6376 GTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTT 6425
            :::   :::|||||||   :::||||||               ||
 1019 ......IleThrPheThrGlnSerAspLysSer............LeuLe 1029

 6426 GAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGG 6475
      |   ::::::|||   |||      |||   |||||||||||||:::||||
 1029 uLeuSerArgGlyTyrGlu...AspValHisThrValHisGluValGlnG 1045

 6476 GAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGAC 6525
      ||:::||||||:::||||||   |||   |||       :::
 1045 lyGluThrPheGluAspValSerLeuValArgLeuThrProThrProVal 1061

 6526 TCCCTATTCACTAAACAA...CCGCATATACTTGTTGGTTTGTCGAGACA 6572
      ::::::::::::||||||   ||||||:::|||||||:::|||||||||||
 1062 GlyIleIleSerLysGlnSerProHisLeuLeuValSerLeuSerArgHi 1078

 6573 CACACGCTCACTGGTTTAT..............GCCGCTCTGAGCTCAA 6607
      ||||||||||:::   |||                  |||:::   |||
 1078 sThrArgSerIleLysTyrTyrThrValValLeuAspAlaValValSerV 1095

 6608 AGTTGGACGATAAG......GTCGGCACATATATTAGCGACGCGTCACCT 6651
        |||   |||         |||::::::|||:::   |||
 1095 alLeuArgAspLeuGluCysValSerSerTyrLeuLeuAspMetTyrLys 1111

 6652 CAATCAGTATCC 6663
         :::||||||
 1112 ValAspValSer 1115
```

seq_name: SwissProt_39:RRPO_PMMVS

seq_documentation_block:

```
ID   RRPO_PMMVS     STANDARD;      PRT;  1612 AA.
AC   P29098; Q84924;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Pepper mild mottle virus (strain Spain) (PMMV-S).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92113528; PubMed=1765765;
RA   Alonso E., Garcia-Luque I., de la Cruz A., Wicke B.,
RA   Avila-Rincon M.J., Serra M.T., Castresana C., Diaz-Ruiz J.R.;
RT   "Nucleotide sequence of the genomic RNA of pepper mild mottle virus,
RT   a resistance-breaking tobamovirus in pepper.";
RL   J. Gen. Virol. 72:2875-2884(1991).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81413; AAB02334.1; -.
DR   EMBL; M81413; AAB02335.1; -.
DR   PIR; JQ1312; WMTMPV.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1612       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1117       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     835    842       ATP (POTENTIAL).
SQ   SEQUENCE   1612 AA;  183265 MW;  465677F4924802BF CRC64;
```

alignment_scores:

Quality: 297.00 Length: 854
Ratio: 0.754 Gaps: 40
Percent Similarity: 46.136 Percent Identity: 22.834

alignment_block:

US-09-301-906-3 x RRPO_PMMVS ..

Align seg 1/1 to: RRPO_PMMVS from: 1 to: 1612

```
 4522 CCCAAACAGAGGAGTTTGACGATC..........................  4545
      ||||||:::::::::::::::|||
  366 ProLysMetLysAspMetValIleValProLeuPheAspValSerLeuGl  382

 4546 .AATGAAGGCAAG......GCCGGTAAACAGCTTTGCATGTTTAGAACGT 4588
       |||||||||||       |||   |||::::::   :::   :::
  382 nAsnGluGlyLysArgLeuAlaArgLysGluValMetValSerLysAspP 399

 4589 GTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACC 4638
              |||          :::   :::      :::   |||
  399 heValTyrThrValLeuAsnHisIleArgThrTyrGlnSerLysAlaLeu 415

 4639 AGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGG..... 4683
         ::::::|||:::::::::|||||::::::::::::::::|||
  416 ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgValIl 432

 4683 ..................................................  4683

  432 eIleAsnGlyValThrAlaArgSerGluTrpAspValAspLysAlaLeuL 449

 4684 ...........AGTGCGGTCTTTTTC......TCAAAGCTGGGT...... 4710
                 |||   :::||||||      :::|||||||:::
  449 euGlnSerLeuSerMetThrPhePheLeuGlnThrLysLeuAlaMetLeu 465

 4711 .................GAGGGGTATACCTATAATGGTGGTAGCCATGT 4742
                        :::   :::      ::::::   |||   ::
  466 LysAspAspLeuValValGlnLysPheGlnValHisSerLysSerLeuTh 482

 4743 TTCATCAGGGTGG.............................CCTC 4759
      ::::      |||                              |||
  482 rGluTyrValTrpAspGluIleThrAlaAlaPheHisAsnCysPheProT 499

 4760 GTGCCCTAGAGGATATCTTAACG..........................  4782
            |||   :::::::::
  499 hrIleLysGluArgLeuIleAsnLysLysLeuIleThrValSerGluLys 515

 4783 ......GCAATTAAGTACCCAAGCGTC.........TTCGACCACTGTTT 4817
            ||||||   |||::::::         |||         ||
  516 AlaLeuGluIleLysValProAspLeuTyrValThrPheHisAspArgLe 532

 4818 AGTGCAGAAGTACAAGATGGGTGGAGGCGTACCA.........TTCCACG 4858
      ||||::::::||||||   :::      :::|||
  532 uValLysGluTyrLysSerSerValGluMetProValLeuAspValLysL 549

 4859 CTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAAT 4908
        :::   ||||||:::            |||   :::      ::::::
```

```
 549 ysSerLeuGluGluAlaGluValMetTyrAsnAlaLeuSerGluIleSer 565

4909 CTCGTGGGGAAGGCAAAC.................TTCTCGACTAAGTG 4940
     ::::::      ::::::                 ||||||      ||
 566 IleLeuLysAspSerAspLysPheAspValAspValPheSerArgMetCy 582

4941 CAGGAAGGGTGGT.....................AAGGTCATGGTCATAA 4969
     |:::      |||                    ||||||||||||
 582 sAsnThrLeuGlyValAspProLeuValAlaAlaLysValMetValAlaV 599

4970 ACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACG 5019
       ||:::
 599 alValSer.......................................... 601

5020 CACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTT 5069
                           :::|||:::   ::::::||||||||
 602 .......................AsnGluSerGlyLeuThrLeuThrPh 610

5070 CAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCG 5119
     |:::         ::::::   :::|||:::      ||||||
 610 eGluArgProThrGluAla...AsnValAlaLeuAlaLeuGln....... 623

5120 GCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAG 5169
                                   |||:::         :::
 624 ...............................ProThrIleThrSerLys 629

5170 AGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCT 5219
     ::::::|||:::   :::   ::::::         |||
 630 GluGluGlySerLeuLysIleValSerSerAspValGly........... 642

5220 ATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATA. 5268
         |||:::|||   :::||||||      :::|||      |||:::
 643 ....GluSerSerIleLysGluValValArgLysSerGluIleSerMetL 658

5269 .................TGGTGCGAACAAGATTACGTTAGGAAGTGTGAG 5301
                           ::::::::::::   |||:::   |||
 658 euGlyLeuThrGlyAsnThrValSerAspGluPheGlnArgSerThrGlu 674

5302 TGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCC 5351
                        :::   :::|||:::
 675 .....................IleGluSerLeuGln.............. 679

5352 AATGACATTTGAAGTGGTTAAAACCGGGACC.........TCTGAAGATG 5392
           |||:::::::|||:::|||   |||  .       ::::::
 680 ....GlnPheHisMetValSerThrGluThrIleIleArgLysGlnMetH 695

5393 CCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATAC 5442
       ::::::   |||         |||   :::         ::::::|||
 695 isAlaMetValTyrThrGlyProLeuLysValGlnGlnCysLysAsnTyr 711

5443 ...AGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGG 5489
           :::|||:::   :::   :::   ||||||::::::   :::
 712 LeuAspSerLeuValAlaSerLeuSerAlaAlaValSerAsnLeuLysLy 728

5490 TGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACG 5539
      ::::::|||
 728 sIleIleLys........................................ 731

5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTG 5589
                   |||      |||      :::|||::::::   :::
 732 ...........AspThrAlaAlaIleAspLeuGluThrLysGluLysPhe 744

5590 CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATC..... 5634
       |||           |||   :::      :::|||   :::
 745 GlyVal............TyrAspValCysLeuLysLysTrpLeuValLy 757

5635 ....TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACA 5680
         ::::::||||||           :::|||:::         |||:
 757 sProLeuSerLysGlyHisAlaTrpGlyValValMet.........AspS 771

5681 ATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTG 5730
     ::   :::   |||         |||:::   :::|||:::   |||:::
 771 erAspTyrLysCysPheValAlaLeuLeuThrTyrAspGlyGluAsnIle 787

5731 ATAGGTGCACTGAAAGTCGCTCGATGCGGC............ATGGTGGG 5768
     :::                    ||||||              |||::
 788 Val.....................CysGlyGluThrTrpArgArgValAl 797

5769 CGAATCGTTTAAGTCCTTCGAATATAAGTGCTAT................ 5802
     :   |||   :::|||:::   |||:::
 797 aValSerSerGluSerLeuValTyrSerAspMetGlyLysIleArgAlaI 814

5803 ....................AATGCTCCCCCAGGTGGCGGTAAGACGACG 5832
                         :::   |||      ::::::      :::
 814 leArgSerValLeuLysAspGlyGluProHisIleSerSerAlaLysVal 830

5833 ACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTAC 5882
     ||||||||||||      |||   :::   ::::::|||      |||
 831 ThrLeuValAspGly...ValProGlyCysGlyLysThrLysGluIleLe 846

5883 GGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAG.... 5928
      ::::::|||:::   :::||||||:::   ::::::      |||
 846 uSerArgValAsnPheAspGluAspLeuValLeuValProGlyLysGlnA 863

5929 .................AGAGATCCGAATTTGGAAGGTCTCAACAGTGCT 5961
                        |||      |||   :::||||||      |||
 863 laAlaGluMetIleArgArgArgAlaAsnSerSerGlyLeuIleValAla 879

5962 ACCACAGTTAACTCCAGGGTGGTA...AACTTTATCGTC.........AG 5999
     |||      |||   |||:::|||   :::|||::::::         ||
 880 ThrLysGluAsnValArgThrValAspSerPheLeuMetAsnTyrGlyAr 896

6000 GGGA......ATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGC 6043
     ||||      :::|||||||||:::::::::||||||      |||:::|
 896 gGlyProCysGlnTyrLysArgLeuPheLeuAspGluGlyLeuMetLeuH 913

6044 ATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGC 6093
     ||   |||   :::::::::   |||      :::   :::||||||:::
 913 isProGlyCysValAsnPheLeuValGlyMetSerLeuCysSerGluAla 929

6094 CTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGT 6143
     :::   :::|||||   :::|||||||||:::|||||||||      ::
 930 PheValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValAlaTh 946

6144 GTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTAT 6193
     :|||                ::::         :::   ::::::|||
 946 rPheProTyrProLysHisLeuSerGlnLeuGluValAspAlaValGluT 963

6194 ACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCC 6243
                  :::   |||||||||   |||:::   ::::::|||:::
 963 hrArgArgThrThrLeuArgCysProAlaAspIleThrPhePheLeuAsn 979

6244 TCA.................ATGACCGTAAGGGGAACGGAAAAGTGTTA 6275
     :::                    |||   :::   ::::::   :::
 980 GlnLysTyrGluGlyGlnValMetCysThrSerSerValThrArgSerVa 996

6276 CCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGT 6325
         ::::::|||:::::::|||            |||::::::   :::|
 996 lSerHisGluValIleGlnGly......AlaAlaValMetAsnProValS 1011

6326 CCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGAC 6375
     |||||         |||   :::
1011 erLysProLeuLysGlyLysVal........................... 1018

6376 GTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTT 6425
           :::   :::||||||   :::||||||                ||
1019 ......IleThrPheThrGlnSerAspLysSer............LeuLe 1029

6426 GAAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGG 6475
     |   ::::::|||   |||      |||   ||||||||||||:::||||
1029 uLeuSerArgGlyTyrGlu...AspValHisThrValHisGluValGlnG 1045

6476 GAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGAC 6525
     ||:::||||||:::||||||   |||   |||         :::
1045 lyGluThrPheGluAspValSerLeuValArgLeuThrProThrProVal 1061
```

```
    6526 TCCCTATTCACTAAACAA...CCGCATATACTTGTTGGTTTGTCGAGACA 6572
         ::::::::::::||||||   ||||||:::||||||:::||||||||||
    1062 GlyIleIleSerLysGlnSerProHisLeuLeuValSerLeuSerArgHi 1078

    6573 CACACGCTCACTGGTTTAT..............GCCGCTCTGAGCTCAA 6607
         ||||||||||:::   |||                 |||:::   |||
    1078 sThrArgSerIleLysTyrTyrThrValValLeuAspAlaValValServ 1095

    6608 AGTTGGACGATAAG......GTCGGCACATATATTAGCGACGCGTCACCT 6651
           |||   |||         |||:::::|||:::   |||
    1095 alLeuArgAspLeuGluCysValSerSerTyrLeuLeuAspMetTyrLys 1111

    6652 CAATCAGTATCC 6663
           :::|||||
    1112 ValAspValSer 1115

seq_name: SwissProt_39:V1A_BBMV

seq_documentation_block:
ID   V1A_BBMV       STANDARD;      PRT;   966 AA.
AC   Q00020;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   1A PROTEIN [INCLUDES: HELICASE; METHYLTRANSFERASE].
OS   Broad bean mottle virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC   Bromovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BAWDEN;
RX   MEDLINE=92074218; PubMed=1962437;
RA   Dzianott A.M., Bujarski J.J.;
RT   "The nucleotide sequence and genome organization of the RNA-1 segment
RT   in two bromoviruses: broad bean mottle virus and cowpea chlorotic
RT   mottle virus.";
RL   Virology 185:553-562(1991).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE REPLICATION OF THE VIRUS.
CC       CONTAINS AN HELICASE DOMAIN AND A METHYLTRANSFERASE DOMAIN. THE
CC       METHYLTRANSFERASE DOMAIN IS PROBABLY INVOLVED IN VIRAL RNA
CC       CAPPING.
CC   -!- SIMILARITY: TO 1A PROTEIN FROM CCMV, CMV, PSV AND TAV.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M65138; AAA42740.1; -.
DR   PIR; A41699; P1BVBB.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Helicase; ATP-binding; Transferase; Methyltransferase.
FT   NP_BIND     690    697       ATP (POTENTIAL).
SQ   SEQUENCE   966 AA;  109621 MW;  DF592681D7231C8D CRC64;


alignment_scores:
            Quality:  290.00            Length:    775
              Ratio:  0.773              Gaps:     33
 Percent Similarity:  48.387   Percent Identity: 23.613

alignment_block:
 US-09-301-906-3 x V1A_BBMV   ..

 Align seg 1/1  to: V1A_BBMV  from: 1  to: 966

    4580 TTAGAACGTGTTCCTGCGGTGTGCAGCTGGATG.....TGTAC 4617
```

```
         ||||||||||:::         ||||||   |||                ::
  275 LeuGluArgMetLeuLeuLysCysSerIleMetThrTyrLysIleIleAl 291

 4618 AACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTT.ACCTTTGTCG 4666
      :|||:::          ||||||   |||      |||      :::   :
  291 aThrAsnValArgCysProProGluSerLeuArgHisCysIleTrpPheG 308

 4667 ATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGG 4716
      ::|||::::::          ||||||         ::::::      |||
  308 luAsnIleSerGlnTyrLeuAlaVal.........GlnIleProIleGly 321

 4717 TATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCT 4766
      |||:::   |||                        |||      :::
  322 TyrAsnLeuAsn.....................AspTrpLysThrValAr 331

 4767 AGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTT 4816
                 |||:::   :::      ::::::      :::|||:
  331 gValAlaArgAlaThrValArgGluValGluGluIleSerPheArgCysP 348

 4817 TAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGAC 4866
      ::   ::: :::      |||:::|||      :::|||:::
  348 heLysGlu.AsnLysAspTrpThrGluAsnMetArgSerValAlaSerIl 364

 4867 GAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGG 4916
          ||||||      :::   :::::: ::::::|||||||||
  364 eLeuSerAlaLysSerSerThrValIle.IleAsnGlyGlnSerIleMet 380

 4917 GAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCA 4966
      :::|||:::   ||||||   ::::::
  381 SerGlyGluArgLeuAspValLeuGlu..................... 389

 4967 TAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGG 5016
                                  :::   :::   ||||||
  390 .......................TyrHisLeuValAlaPheSerLeuT 398

 5017 ACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGAC 5066
               :::::::|||   ::::::   |||      |||   :::
  398 hrLeuAsnLeuTyrGlnLysTyrGluLysLeuArgAsnPheGlnGlyGlu 414

 5067 GTTCAGGGCAAC.................TCGGCGCGTCTT...TGGTG 5095
      ::::::   :::                 ::::::||||||   |||||
  415 LeuGluTrpLysGlyTrpAlaAsnHisPheLysThrArgLeuTrpTrpCy 431

 5096 TAGGCAGGATGTTGCAGTTAGCCGGCGGC..................... 5124
      | ||||||   ::::::      ||||||
  431 s.GlyArgThrValSerThrGluGlyGlyPheLeuArgAsnPheLeuAla 447

 5124 ..................................................... 5124

  448 AspLysIleProTrpLeuLysLeuAsnThrTyrAlaAspSerLeuAspPh 464

 5125 ..........GTGTCGGATGAGAAGTCA.........CCAGGTGTTCCAA 5155
                ::::|||:::   :::|||          :::|||||||:
  464 eIleThrLysIleSerGluValGluSerPheGluValAspSerValProT 481

 5156 ACCAGCAACCACAGAGC...CAAGGTGCTACCAGAACAATCACACCAAAA 5202
      ::::::::   :::|||               ::::::|||:::   :::
  481 hrSerArgLeuArgSerPhePheGlnLysGluGluAsnIleValGluArg 497

 5203 TCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAG 5252
      ::::::::::::   :::|||   ::::::   |||   :::|||
  498 AlaAlaSerGluIleMetSerAlaAsnAlaArgArgIleAlaLys..... 512

 5253 GTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGT 5302
                                                |||:::|||
  513 ....................................LysAlaGluM 516

 5303 GGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCA 5352
           :::       :::       :::             ::::::::::|||
  516 etSerLysGluPheAspAspPheValAspAlaProGluGluPheAlaPro 532

 5353 ......ATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGT 5396
            ::::::   ||||||::::::|||   ::::::::::|||:::
```

```
  533 GluAspValValGluGluValIleAsnThrProValThrGlnAspValLy 549

 5397 CGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGG 5446
      ::::::     |||                    :::|||   |||:
  549 sLeuArgGlnSerLys.................ProGluThrAlaArgS 560

 5447 CGTTGCTTATG..............GCTAGAAATATTGCCGTCAC.... 5477
      ::::::::::                ::::||  .|||:::
  560 erIleValLeuAspProAspAlaValLeuLysAsnGlyAlaIleAsnGlu 576

 5478 .........TACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATG 5518
              |||   :::|||      |||:::|||   |||:::||| :
  577 PheAlaAspTyrSerLysArgLeuHisGluAsnThrValSerAsnLeu.A 593

 5519 AATCACTA.................CCGGGCTTTCACGTTTACAAGTCG 5550
      ::   |||                    |||   ::::::::::::::
  593 rgHisLeuTrpThrLeuMetGlyCysArgGlyAsnGluIleHisAsnLys 609

 5551 GGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGA.. 5598
      ::::::          :::|||   ::::::|||   ::::::|||:::
  610 SerValAlaGluThrTyrHisArgValAspAspMetValAsnValHisPh 626

 5599 ......................GACCTACCGTACGTACTCATAGCTGAAA 5626
                            |||   |||         :::
  626 eProAsnGlyHisTrpMetTyrProLeuLysTyrGluTyrThrValGlyT 643

 5627 AAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGC 5676
        :::          |||:::      :::      :::   :::
  643 yrAsnAspGlyGlyLeuGlyGluLysPheGluAsnGluLeuTyrValVal 659

 5677 GACAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAA 5726
      |||:::          |||   :::      ::::::   ||||||   ::
  660 AspLysThrCysSerCysAlaAsnAlaLysAlaIleAlaAspAlaCysLy 676

 5727 TTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGT 5776
      :   ::::::|||               |||::::::|||
  676 sLysValSerAla.........ProThrCysSerValVal.......... 686

 5777 TTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAG 5826
                              ::::::     |||   ||||||
  687 ...................MetValAspGlyValAlaGlyCysGlyLys 696

 5827 ACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCAC 5876
      ||||||::::::   :::   |||           ::::::
  697 ThrThrAlaIleLysGluThrPhe......ArgPheGluLysAspIleIl 711

 5877 CATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGA 5926
       :::|||||||||      |||:::||||||::::::   |||:::
  711 eValThrAlaAsnArgLysSerAlaGluAspValArgLysAlaIle.... 726

 5927 AGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCC 5976
                          |||      |||   :::   |||
  727 ................PheGlyAspAlaSerAspSerGluValAlaLeu 737

 5977 AGGGTGGTA............AACTTTATCGTCAGGGGAATGTATAAA.. 6012
      :::||||||           :::   |||::::::|||:::   :::
  738 LysValValArgThrAlaAspSerAlaIleMetHisGlyLeuProGluCy 754

 6013 ....AGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTAC 6058
          |||:::|||||||||||||:::   ::::::|||   |||   |||
  754 sHisArgLeuLeuValAspGluAlaGlyLeuLeuHisTyrGlyGlnLeuL 771

 6059 AACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGAC 6108
              :::          :::||||||   |||   |||||||||
  771 euAlaValAlaAspLeuCysLysCysSerGluValLeuAlaPheGlyAsp 787

 6109 ATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGAT.. 6156
        :::||||||   |||   :::|||:::   :::|||||||||
  788 ThrGluGlnIleSerPheLysSerArgAspAlaThrPheArgMetLysTy 804

 6157 .TGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAAT 6205
       |||   :::         |||::::::   |||         ||||||:
  804 rCysAsnIleGluTyrAspLysArgAspIleVal.........SerLysT 818

 6206 CGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTA 6255
      :::::|||||||||   ||||||         :::|||::::::   :::
  818 hrPheArgCysProGlnAspVal.........ValSerAlaValLysIle 831

 6256 AGGGGAACGGAAAAGTGT.................TACCCTGAAAAGGT 6287
                ::::||||||                |||         ||
  832 Leu...LysArgLysCysAlaAsnArgSerSerLysTyrAsnGlyTrpVa 847

 6288 CGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAA 6337
      |   |||:::|||         |||   ::::::::::::::      |
  847 lSerSerSerLys.........ValGluLysSerLeuSerLysSerArgI 861

 6338 TTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGC 6387
      ||   :::   ::::::|||:::      :::         ||||||
  861 leValSerIleAsnGlnValSerMetGluLysHisLysPheTyrLeuThr 877

 6388 ATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGG 6437
      ||||||:::   :::|||:::   :::         |||     ||
  878 MetThrGluAlaAspLysAlaAlaLeuCysSerArgAlaLysAspValGl 894

 6438 A..............AAAGAAACACCAGTGATGACAGTGCATGAAGCAC 6472
      |                  |||:::   :::   ||||||||||||||||
  894 yLeuAspLysThrTrpValGluSerAsnMetGluThrValHisGluAlaG 911

 6473 AGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGAT 6522
      ||||||||:::   :::   |||||||||   |||   |||::::::
  911 lnGlyLysAlaValAspHisValValLeuValArgLeuLysSerThrLys 927

 6523 GACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACA 6572
         :::||||||   ::::::   :::   ||||||:::|||:::||||||
  928 CysAspLeuPheLysSerGluGluTyrCysLeuValAlaLeuThrArgHi 944

 6573 CACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGG 6622
      |   |||::::::   |||         ::::::         |||   :
  944 sLysArgThrPheGluTyrLeuTyrAsnGlyAspLeuGlyGlyAspLeuI 961

 6623 TCGGCACATATATTAGC 6639
      :::::   |||::::::
  961 leSerPheTyrValThr 966

seq_name: SwissProt_39:RRPO_TMVRA

seq_documentation_block:
ID   RRPO_TMVRA     STANDARD;      PRT;  1616 AA.
AC   Q98745; O41340;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tobacco mosaic virus (strain Rakkyo) (TMV-R).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96265021; PubMed=8678834;
RA   Chen J., Watanabe Y., Sako N., Ohsima K., Okada Y.;
RT   "Complete nucleotide sequence and synthesis of infectious in vitro
RT   transcripts from a full-length cDNA clone of a rakkyo strain of
RT   tobacco mosaic virus.";
RL   Arch. Virol. 141:885-900(1996).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; D63809; BAA09876.1; -.
DR   EMBL; D63809; BAA09877.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183579 MW;  AC7CF5429EBAD49A CRC64;


alignment_scores:
           Quality:  288.00            Length:   1880
             Ratio:  0.425               Gaps:     62
 Percent Similarity:  36.064   Percent Identity: 16.809

alignment_block:
 US-09-301-906-3 x RRPO_TMVRA   ..

 Align seg 1/1  to: RRPO_TMVRA  from: 1  to: 1616

    1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
         |||   :::|||       :::   |||:::::::::::::::::::   ::
       2 AlaTyrThrGlnThrAlaThrThrSerAlaLeuLeuAspThrValArgGl 18

    1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
         ::::      |||   ::::::   |||               ||||||
      18 yAsnAsnSerLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

    1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
           :::   |||   |||:::   ||||||            ::::::|||
      35 laVal...AspGluPheAsnAlaArgAsp.............ArgArgPro 46

    1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
         :::   :::   :::      ::::::   ::::::
      47 LysValAsnPheSerLysValIleSerGluGluGlnThrLeuIleAlaTh 63

    1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
          :::   :::|||||:::::::|||   |||   ::::::   ::::::
      63 rArgAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

    1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
           |||   :::|||:::::::::||||||      ::::::   ::::::
      80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

    1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
           :::         :::         :::|||||||||||||:::|||::
      97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

    1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
         :   |||:::      |||::::::      |||   |||   |||   |
     113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

    1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.............ATC 1641
         ·   ||||||||:::|||:::   |||:::   ::::::
     130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

    1642 CTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCT 1691
         |||:::||||||               :::   :::      ::::::
     147 LeuTyrLeuSerArgLeuAspArgGlyGlyLysThrValProAsnPheGl 163

    1692 AACTGAAGCGGCTTCAAAGTCTGTGTCTTAC..............TGTA 1726
              ||||||   ::::::   :::::::::              |||
     163 nLysGluAlaPheAspArgTyrAlaGluPheProGluAspAlaValCysH 180

    1727 GTCGAGAATCGCAGAACTGCGAT...................TCTAGA 1755

         :::      |||:::|||:::                      ::::::
     180 isAsnThrPheGlnThrCysGluHisGlnProMetGlnGlnSerGlyLys 196

    1756 GCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGT 1805
         :::   |||       :::   ::::::|||||||||      ::::::
     197 ValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAspGluPh 213

    1806 AGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTA 1855
          :::   |||:::   :::|||::::::   :::         |||:::
     213 eGlyAlaAlaLeuLeuArgLysAsnValHisThrCysTyrAlaAlaPheH 230

    1856 TGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAA 1905
           |||       :::||||||      ::::::   |||   |||:::|||
     230 isPheSerGluAsnLeuLeuLeuGluAspSerTyrValAsnLeuAspGlu 246

    1906 CTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGG 1955
         :::::::::       :::|||:::||||||   |||::::::::::   ::
     247 IleAsnAlaCysPheSerArgAspGlyAspLysLeuThrPheSerPheAl 263

    1956 TCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAAGCGGGT 2002
         ::::   :::   :::   |||   |||||||||||||||:::      :
     263 aSerGluSerThrLeuAsnTyrCysHisSerPheSerAsnIleLeuLysT 280

    2003 TTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTA 2052
         ::         :::|||      :::|||:::   :::|||:::
     280 yrValCysLysThrTyrPheProAlaSerAsnArgGluValTyrMetLys 296

    2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
         |||:::
     297 GluPhe..................................... 298

    2103 TCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGT 2152

     298 ..................................................  298

    2153 TCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTG 2202

     298 ..................................................  298

    2203 TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTC 2252
                                          :::|||   |||:::   ::
     299 ..........................LeuValThrArgValAsnTh 305

    2253 CTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTG 2296
         ::::         ::::::   |||   ||||||
     305 rTrpPheCysLysPheSerArgIleAspThrPhe................ 316

    2297 AGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGG 2346
                                                      :::
     317 ....................................LeuLeuTyr 319

    2347 AGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGG 2396
         :::   |||:::|||
     320 LysGlyValAlaHis.............................. 324

    2397 TTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTA 2446
                                                    ||||||:
     325 ........................................LysSerV 327

    2447 TTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTC 2496
         ::   |||   :::|||       |||   |||   :::
     327 alAspSerGluGlnPheTyrThrAlaMetGluAspAlaTrpHis...... 341

    2497 TTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTC 2546
                 :::   :::|||:::::   |||   |||:::   :::
     342 .........TyrLysLysThrLeuAlaMetCysAsnSerGluArgIleLe 355

    2547 CGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGC 2596
          :::      ::::::   |||:::      |||      ::::::   :
     355 uLeuGluAspSerSerSerValAsnTyrTrpProLysMetArgAspM 372

    2597 TTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTT 2646
         :::::   :::|||
```

```
 372 etValIleIlePro..................................... 376

2647 GTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGC 2696

 377 .............................................Le 377

2697 CTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCT 2746
      ||||||||||||   ||||||              |||:::||||||
 377 uPheAspIleSerLeuGluThr............SerLysArgSerArgL 390

2747 TGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATT...AAGGTCATG 2793
       :::|||:::         :::   :::|||   :::      |||:::
 390 ysGluValLeu.........ValSerLysAspPheValPheThrValLeu 403

2794 AAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTC 2843
     :::   ::::::|||   ::::::|||   |||      |||      ::
 404 AsnHisIleArgThrTyrGlnAlaLysAlaLeu......ThrTyrAlaAs 418

2844 TGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGG 2893
     :|||   :::|||:::   :::      |||:::|||
 418 nValLeuSerPheValGluSerIleArgSerArgVal............. 430

2894 CAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAG 2943
          :::::::::|||      |||            :::|||:::|||
 431 .....IleIleAsnGlyValThrAlaArgSerGluTrpAspValAspLys 445

2944 GGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTC 2993
     :::|||:::         :::   |||
 446 SerLeuLeuGlnSerLeuSerMetThr..................... 454

2994 GTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATG 3043
      ||||||
 455 .PhePhe..................................... 456

3044 AAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTG 3093
                                        :::::::::   :::
 457 ..............................LeuHisThrLysLeu 461

3094 GGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAAC 3143
     ::::::|||:::   :::   :::   :::||||||         |||::
 462 AlaValLeuLysAspAsp...LeuLeuIleSerLysPheSerLeuGlySe 477

3144 GAAA............GAATTTTGGAGTGAAGTTTCCTTA.........A 3172
     :|||           :::   |||:::|||:::||||||         |
 477 rLysThrValCysGlnHisValTrpAspGluIleSerLeuAlaPheGlyA 494

3173 ATGACTTCACCACTTTCGTA......TTGCGGAATAAGGTGCTTATCGGG 3216
     ||   |||   :::            |||   ||||||   ||||||
 494 snAlaPheProSerValLysGluArgLeuLeuAsnLysLysLeuIle... 509

3217 ATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCG 3266

 509 ..................................................... 509

3267 CGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTC 3316
                          |||   ||||||::::::   |||
 510 ......................ArgAlaAlaGlyAspAlaLeuGluIle. 518

3317 TAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCT 3366

 518 ..................................................... 518

3367 AGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGC 3416

 518 ..................................................... 518

3417 GGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGT 3466
                                        :::||||||
 519 ................................LysValProAspLeuT 524

3467 TGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGC 3516
       ::::::|||
 524 yrIleThrPhe..................................... 527

3517 ATTTGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGG 3566

 527 ..................................................... 527

3567 GACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCT 3616

 527 ..................................................... 527

3617 TCTCCAATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGT 3666

 527 ..................................................... 527

3667 GTCAGCTCATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAG 3716

 527 ..................................................... 527

3717 GGGCTTGCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAG 3766

 527 ..................................................... 527

3767 AGGGGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGAC 3816
                    :::|||:::   :::      |||:::
 528 ..............HisAspArgLeuValAlaGluTyrLys......... 536

3817 TATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGG 3866

 536 ..................................................... 536

3867 TGACGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCA 3916
                       |||:::   ::::::   :::   |||:::
 537 ................SerSerValAspMetProAlaLeuAspIleArgL 548

3917 CAAATGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAA 3966
       ::::::
 548 ysArgMet..................................... 550

3967 GAAGAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGA 4016
     |||||||||            :::|||:::|||:::   ::::::|||
 551 GluGluThr............GluValMetTyrAsnAlaLeuSerGlu.. 562

4017 GGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTT 4066

 562 ..................................................... 562

4067 CTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTC 4116
                          ::::::   :::::::::      |||
 563 ................LeuSerValLeuArgGluSerAspLysPheAsp 573

4117 GACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGA 4166
        |||||||||   |||:::                ::::::::::::||
 574 ValAspValPheSerGlnMet..............CysLysSerLeuGl 585

4167 GGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGA 4216
     |||||||         :::         :::|||:::      :::   :
 585 uValAsp......ProMetThrAlaAlaLysValIleValAlaValMetS 600

4217 CACAAGAAAGGGGA..................AGAGCTGCAGATGGTAAC 4248
     :::::|||   |||                |||   :::::::::|||
 600 erAsnGluSerGlyLeuThrLeuThrPheGluArgProThrGluAlaAsn 616

4249 ACAGCACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAG 4298
     :::||||||   |||:::::::::   ::::::            :::
 617 ValAlaGln...AlaLeuGlnAspGlnGluLysAlaSerGluGlyAlaLe 632

4299 TCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGG 4348
      :::|||:::|||   :::::::::   |||::::::      |||
 632 uValValThrSerArgGluValGluGluProSerMetLysGlySer.... 647

4349 TACATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCG 4398
                       ::::::   |||:::   ||||||      :::
 648 ..............MetAlaArgGlyGluLeuGlnLeuAlaGlyPheAla 659
```

```
4399 GGCGCCACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCG 4448
     |||      |||
 660 GlyAspHisPro..................................... 663

4449 TGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGA 4498
                                          :::   :::   :
 664 ..................................GluSerSerTyrS 668

4499 AAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAAT 4548
     ::::::::  :::::::|||:::   :::|||      ::::::
 668 erArgAsnGluGluIleGluSerLeuGluGlnPheHisMetAlaThrAla 684

4549 GAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGG 4598
     ::::::         ||||||:::      :::
 685 AspSerLeuIleArgLysGlnMetSerSerIle................. 695

4599 TGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAA 4648
                 |||||||:::          :::   :::::::::|
 696 .............ValTyrThrGlyProIleLysValGlnGlnMetLysA 708

4649 ACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTC 4698
     ||         |||:::|||:::|||   :::      ::::::
 708 sn.........PheIleAspSerLeuValAlaSerLeuSerAla...... 719

4699 TCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATC 4748

 719 ..................................................... 719

4749 AGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAA 4798
                 |||:::::::::::::::      :::|||      :
 720 .............AlaValSerAsnLeuValLysIleLeuLysAspThrA 732

4799 GCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTA 4848
     :::::::|||      |||      :::   |||:::|||
 732 laAlaIleAsp......LeuGluThrArgGlnLysPheGly......... 743

4849 CCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTT 4898

 743 ..................................................... 743

4899 GACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGG 4948

 743 ..................................................... 743

4949 GTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATG 4998
                 |||:::::::|||||||||      :::::::::
 744 .............ValLeuAspValAlaSerArgLysTrpLeuIleLys 755

4999 CCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGA 5048
     |||
 756 Pro.................................................. 756

5049 AGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAG 5098

 756 ..................................................... 756

5099 GCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGT 5148

 756 ..................................................... 756

5149 GTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACC 5198
                                                    |||
 757 ..........................................ThrAl 758

5199 AAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGG 5248
      ||||||   :::   ::::::   |||      :::|||:::
 758 aLysSerHisAlaTrpGlyValValGluThrHisAlaArgLys....... 772

5249 GGAGGTCGACATACTCGATA............TGGTGCGAACAAGATTAC 5286
                |||   :::           :::   ||||||
 773 ...........TyrHisValAlaLeuLeuGluTyrAspGluGlnGlyIle 785

5287 GTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGA 5336
```

```
     |||         :::|||   |||
 786 ValThrCysAspAspTrpArgArg.......................... 793

5337 ACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTG 5386
                                          :::::::::|||:
 794 ..................................ValAlaValSerS 798

5387 AAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGG 5436
     ::::::::||||||            |||      ::::::      |||
 798 erGluSerValVal............TyrSerAspMetAlaLysLeuArg 810

5437 ACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGA 5486
     |||   |||   ||||||         :::      |||:::::::|||::
 811 ThrLeuArgArgLeuLeuArgAspGlyGluProHisValSerAsnAlaLy 827

5487 AGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTC 5536
     :   ||||||
 827 sValValLeu.......................................... 830

5537 ACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGC 5586

 830 ..................................................... 830

5587 TTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTT 5636

 830 ..................................................... 830

5637 TACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGT 5686
                    ||||||:::|||
 831 ................ValAspGlyVal...................... 834

5687 TCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGT 5736

 834 ..................................................... 834

5737 GCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTT 5786

 834 ..................................................... 834

5787 CGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGT 5836
                                  ||||||   ||||||
 835 ........................ProGlyCysGlyLys.......... 839

5837 TAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCT 5886
                                       |||      |||   :::
 840 ..............................ThrLysGluIleLeuSer 845

5887 AATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAG........ 5928
     :::|||:::   :::|||||||:::   ::::::      |||
 846 ArgValAsnPheAspGluAspLeuIleLeuValProGlyLysGlnAlaAl 862

5929 .............AGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACC. 5964
                    |||      |||   :::|||:::      ||||||
 862 aGluMetIleArgArgArgAlaAsnSerSerGlyIleIleValAlaThrL 879

5965 ..............ACAGTTAACTCCAGGGTGGTAAACTTT......ATC 5994
                   ||||||:::|||   ::::::||||||
 879 ysAspAsnValArgThrValAspSerPheMetMetAsnPheGlyLysThr 895

5995 GTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCA 6044
     :::|||   ::::::||||||:::::::::||||||      |||:::||
 896 ThrArgCysGlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHi 912

6045 TCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCC 6094
     |   |||   :::::::::   |||      :::   ::::::   :::
 912 sThrGlyCysValAsnPheLeuValAlaMetSerLeuCysAspValAlaT 929

6095 TCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTG 6144
          :::||||||   :::|||||||||:::|||||||||   :::
 929 yrValTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValSerGly 945

6145 TTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATA 6194
     |||         :::                :::   ::::::|||
```

```
 946 PheProTyrProAlaHisPheSerLysLeuGluValAspGluValGluTh 962

6195 CACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCT 6244
              :::   |||||||||   ||||||   :::   |||:::
 962 rArgArgThrThrLeuArgCysProAlaAspValThrHisTyrLeuAsnA 979

6245 CA.................ATGACCGTAAGGGGAACGGAAAAGTGTTAC 6276
                         :::::::::   :::::::::|||
 979 rgArgTyrGluGlyPheValValSerThrSerSerValLysLysServal 995

6277 CCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTC 6326
        :::::::::|||||||||           |||::::::   :::||
 996 SerGlnGluMetValSerGly......AlaAlaValIleAsnProIleSe 1010

6327 CAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACG 6376
     ||||         |||
1010 rLysProLeuHisGlyLys............................... 1016

6377 TGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTG 6426
          |||   :::||||||         ||||||   |||:::::|||
1017 ..IleLeuThrPheThrGln.........SerAsp...LysGluAlaLeu 1028

6427 AAGGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGG 6476
        ::::::|||      :::   |||   |||||||||||:::||||
1029 LeuSerArgGly...TyrSerGluValHisThrValHisGluValGlnGl 1044

6477 AAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACT 6526
     |:::|||:::|||||||||   |||   |||         :::      |
1044 yGluThrTyrSerAspValSerLeuValArgLeuThrProThrProIleS 1061

6527 CCCTATTCACT...AAACAACCGCATATACTTGTTGGTTTGTCGAGACAC 6573
     ||:::::::::      :::||||||:::||||||:::||||||||||
1061 erIleIleAlaGlyAspSerProHisValLeuValAlaLeuSerArgHis 1077

6574 ACACGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAA 6608
     |||   ||||||   |||   :::               |||   |||
1078 ThrCysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIl 1094

6609 GTTGGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
       :::   |||      |||::::::::::|||:::   |||
1094 eIleArgAspLeuGluLysLeuSerSerTyrLeuLeuAsp 1107
```

seq_name: SwissProt_39:RRPO_TMVB

seq_documentation_block:

```
ID   RRPO_TMVB      STANDARD;      PRT;  1616 AA.
AC   Q9YQ08; Q9WKD4;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tobacco mosaic virus (strain B935A) (TMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xue C., Zhou X., Chen Q., Qi Y., Li D.;
RT   "Complete nucleotide sequence and genome organization of tobacco
RT   mosaic virus isolated from Vicia faba.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011933; CAA09875.1; -.
DR   EMBL; AJ011933; CAA09874.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183296 MW;  99A63B441290E6CF CRC64;
```

alignment_scores:

```
            Quality:  285.00              Length:   1879
              Ratio:  0.421                 Gaps:     62
 Percent Similarity:  36.030   Percent Identity: 16.924
```

alignment_block:

US-09-301-906-3 x RRPO_TMVB ..

Align seg 1/1 to: RRPO_TMVB from: 1 to: 1616

```
1207 GCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACCAA 1256
     |||   :::|||      :::   |||:::::::::::::::::   ::
   2 AlaTyrThrGlnThrAlaThrThrSerAlaLeuLeuAspThrValArgGl 18

1257 TAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGA 1306
     ::::      |||   ::::::   |||               ||||||
  18 yAsnAsnSerLeuValAsnAspLeuAlaLysArgArgLeuTyrAspThrA 35

1307 AATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCC 1356
       :::   :::   |||:::   ||||||           ::::::|||
  35 laVal...GluGluPheAsnAlaArgAsp............ArgArgPro 46

1357 TCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAG 1406
     :::   :::   :::       ::::::   ::::::
  47 LysValAsnPheSerLysValIleSerGluGluGlnThrLeuIleAlaTh 63

1407 GGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTA 1456
      :::   :::||||||::::::|||   |||   ::::::   ::::::
  63 rArgAlaTyrProGluPheGlnIleThrPheTyrAsnThrGlnAsnAlaV 80

1457 GTCACCCATTCGCTAATGCCATGCGGAGCTGT...TTCAATGGAATCTTT 1503
       |||   :::|||::::::::::||||||       ::::::   ::::::
  80 alHisSerLeuAlaGlyGlyLeuArgSerLeuGluLeuGluTyrLeuMet 96

1504 TCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCAC 1553
        :::         :::         :::||||||||||||:::|||::
  97 MetGlnIleProTyrGlySerLeuThrTyrAspIleGlyGlyAsnPheAl 113

1554 GTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCC 1603
     :   |||:::      |||::::::      |||   |||   |||   |
 113 aSerHisLeuPheLysGlyArgAlaTyrValHisCysCysMetProAsnL 130

1604 TAGACGTTAAAGATGTGAAGCGGAGAATCAATGAG.....ATC 1641
     ||||||||:::|||:::   |||:::   ::::::
 130 euAspValArgAspIleMetArgHisGluGlyGlnLysAspSerIleGlu 146

1642 CTCTTTCTTTCC......ACAGCTGGGGGAGATTCGTAC..... 1674
     |||:::||||||           ||||||   :::
 147 LeuTyrLeuSerArgLeuGluArgGlyGlyLysThrValProAsnPheGl 163

1675 ....GTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTT 1720
           :::   |||      :::|||      :::   :::|||
 163 nLysGluAlaPheAspArgTyrAlaGluIleProGluAspAlaVal.... 178

1721 ACTGTAGTCGAGAATCGCAGAACTGCGAT..................... 1749
       |||   :::      |||:::|||:::
```

```
 179 ..CysHisAsnThrPheGlnThrCysGluHisGlnProMetGlnGlnSer 194

1750 TCTAGAGCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCA 1799
     :::|||:::   |||      :::   ::::::||||||||      ::
 195 GlyArgValTyrAlaIleAlaLeuHisSerIleTyrAspIleProAlaAs 211

1800 GCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAG 1849
     ::::   :::   |||:::   :::|||:::::   :::          |
 211 pGluPheGlyAlaAlaLeuLeuArgLysAsnValHisThrCysTyrAlaA 228

1850 CTCTTATGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTG 1899
     ||:::   |||      :::||||||      ::::::   |||   |||
 228 laPheHisPheSerGluAsnLeuLeuLeuGluAspSerTyrValAsnLeu 244

1900 GAAGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAA 1949
     :::|||:::::::::      :::|||:::||||||   |||:::::::
 245 AspGluIleAsnAlaCysPheSerArgAspGlyAspLysLeuThrPheSe 261

1950 TGTTGGTCAGTGTGGTGAGATG...TATGAACATTCCTTCTCTAACGTAA 1996
     :   ::::::   :::   :::   |||   ||||||:::||||||:::
 261 rPheAlaSerGluSerThrLeuAsnTyrCysHisSerTyrSerAsnIleL 278

1997 GCGGGTTTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTT 2046
        :::         :::|||      :::|||:::   :::|||:::
 278 euLysTyrValCysLysThrTyrProAlaSerAsnArgGluValTyr 294

2047 AAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTG 2096
           |||:::
 295 MetLysGluPhe..................................... 298

2097 TAGGGCTCAGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCC 2146

 298 ..................................................... 298

2147 CGTCGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCT 2196

 298 ..................................................... 298

2197 AGTGTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGAT 2246
                                        :::|||   |||::
 299 .................................LeuValThrArgVa 303

2247 CTATTCCTAC......GCGCTCAACACTATAGGGACATTCGAGAATAGAA 2290
     :   ::::::         ::::::   |||   ||||||
 303 lAsnThrTrpPheCysLysPheSerArgIleAspThrPhe..... 316

2291 CGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATT 2340
                                                    :::
 317 .............................................Leu 317

2341 ACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATAT 2390
        :::   |||:::|||
 318 LeuTyrLysGlyValAlaHis..... 324

2391 GTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGA 2440
                                                    |
 325 .............................................L 325

2441 AGAGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCC 2490
     |||||:::   |||   :::|||     . |||   |||   :::
 325 ysSerValAspSerGluGlnPheTyrThrAlaMetGluAspAlaTrpHis 341

2491 GGGGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGC 2540
                  :::   :::|||:::::   |||   |||:::
 342 ..............TyrLysLysThrLeuAlaMetCysAsnSerGluAr 353

2541 AGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGG 2590
      :::   :::      ::::::   |||:::      |||      ::::
 353 gIleLeuLeuGluAspSerSerSerValAsnTyrTrpPheProLysMetA 370

2591 AGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCT 2640
     ::   ::::::   :::|||
 370 rgAspMetValIleValPro..... 376

2641 GTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAG 2690

 376 ..................................................... 376

2691 GGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAAT. 2739
            |||||||||||!   ||||||      |||      |||:::
 377 ....LeuPheAspIleSerLeuGluThrSerLysArgThrArgLysGluV 392

2740 ........AGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCA 2781
             |||:::      ||||||:::
 392 alLeuValSerLysAspPheValPheThr..... 401

2782 ATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAAT 2831
          |||:::::::   :::::::|||   :::::::|||   |||
 402 ......ValLeuAsnHisIleArgThrTyrGlnAlaLysAlaLeu..... 414

2832 TACTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCC 2881
      |||      :::|||   :::|||:::   :::      |||:::|||
 415 .ThrTyrAlaAsnValLeuSerPheValGluSerIleArgSerArgVal. 430

2882 ACTGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCG 2931
                        ::::::::::|||      |||
 431 ................IleIleAsnGlyValThrAlaArgSerGluTrp 441

2932 AGTGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTC 2981
     :::|||:::|||:::|||:::
 442 AspValAspLysSerLeuLeuGln..... 449

2982 TGGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTT 3031
                         |||   :::      ||||||
 450 ..................SerLeuSerMetThrPheTyr..... 456

3032 TAACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATA 3081
                                                    :::
 457 .............................................Leu 457

3082 AACTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGT 3131
     ::::::   ::::::::::|||:::   :::   :::   :::||||||
 458 HisThrLysLeuAlaValLeuLysAspAsp...LeuLeuIleSerLysPh 473

3132 CGTAGCTGGAACGAAA.............GAATTTTGGAGTGAAGTTTCCT 3169
           |||:::|||            :::   |||:::|||:::||||
 473 eSerLeuGlySerLysThrValCysGlnHisValTrpAspGluIleSerL 490

3170 TAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATA 3219
     ||
 490 eu..... 490

3220 TTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGG 3269

 490 ..................................................... 490

3270 AATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAA 3319

 490 ..................................................... 490

3320 GCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGC 3369

 490 ..................................................... 490

3370 ACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGT 3419

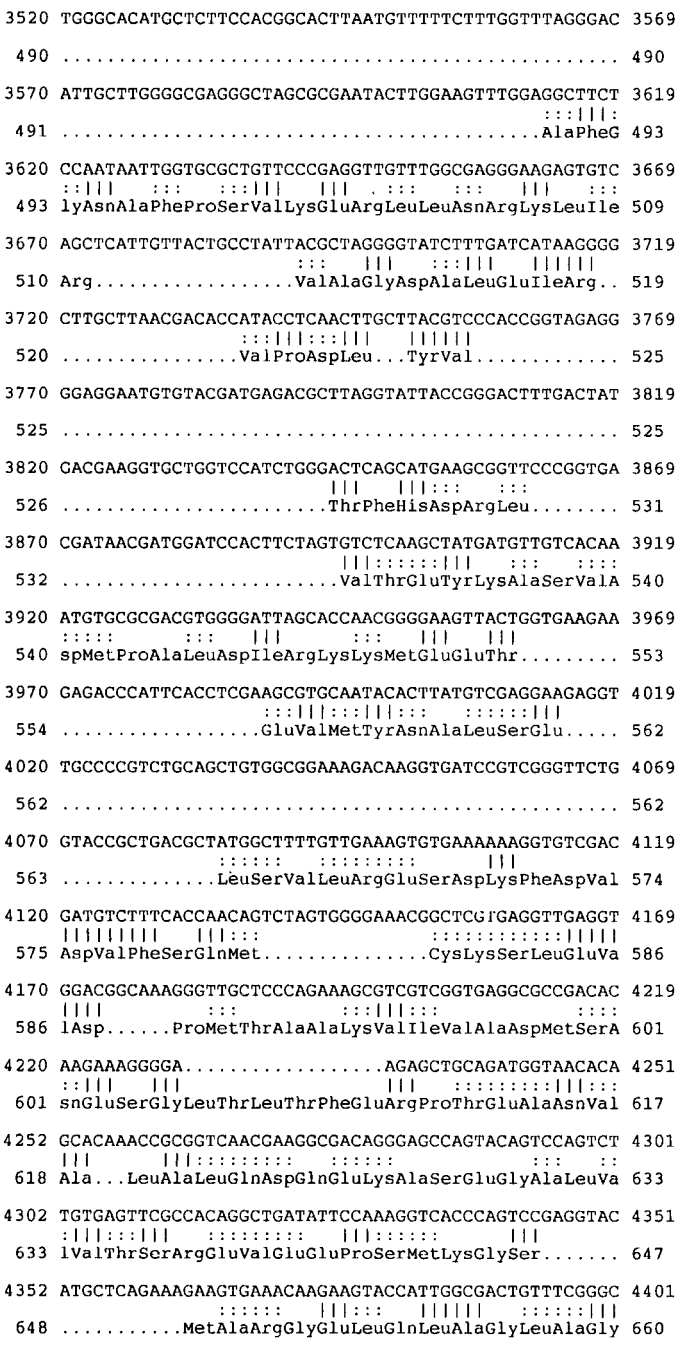
 490 ..................................................... 490

3420 GACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGC 3469

 490 ..................................................... 490

3470 TATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATT 3519

 490 ..................................................... 490
```


3520 TGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGAC 3569
490 490
3570 ATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCT 3619
491 AlaPheG 493
3620 CCAATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTC 3669
493 lyAsnAlaPheProSerValLysGluArgLeuLeuAsnArgLysLeuIle 509
3670 AGCTCATTGTTACTGCCTATTACGCTAGGGGTATCTTTGATCATAAGGGG 3719
510 Arg.....ValAlaGlyAspAlaLeuGluIleArg.. 519
3720 CTTGCTTAACGACACCATACCTCAACTTGCTTACGTCCCACCGGTAGAGG 3769
520 ValProAspLeu...TyrVal..... 525
3770 GGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTGACTAT 3819
525 525
3820 GACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGA 3869
526 ThrPheHisAspArgLeu..... 531
3870 CGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAA 3919
532 ValThrGluTyrLysAlaSerValA 540
3920 ATGTGCGCGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAA 3969
540 spMetProAlaLeuAspIleArgLysLysMetGluGluThr..... 553
3970 GAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGT 4019
554 GluValMetTyrAsnAlaLeuSerGlu..... 562
4020 TGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTG 4069
562 562
4070 GTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGAC 4119
563 LeuSerValLeuArgGluSerAspLysPheAspVal 574
4120 GATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGT 4169
575 AspValPheSerGlnMet.....CysLysSerLeuGluVa 586
4170 GGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACAC 4219
586 lAsp.....ProMetThrAlaAlaLysValIleValAlaAspMetSerA 601
4220 AAGAAAGGGGA.....AGAGCTGCAGATGGTAACACA 4251
601 snGluSerGlyLeuThrLeuThrPheGluArgProThrGluAlaAsnVal 617
4252 GCACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCT 4301
618 Ala...LeuAlaLeuGlnAspGlnGluLysAlaSerGluGlyAlaLeuVa 633
4302 TGTGAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTAC 4351
633 lValThrSerArgGluValGluGluProSerMetLysGlySer..... 647
4352 ATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGC 4401
648 MetAlaArgGlyGluLeuGlnLeuAlaGlyLeuAlaGly 660
4402 GCCACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGG 4451
661 AspHisPro..... 663
4452 TGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAA 4501
664 GluSerSerTyrSerL 669
4502 AACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAA 4551
669 ysAsnGluGluIleGluSerLeuGluGlnPheHisMetAlaThrAlaAsp 685
4552 GGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGT 4601
686 SerLeuIleArgLysGlnMetSerSerIle..... 695
4602 GCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACG 4651
696 ValTyrThrGlyProIleLysValGlnGlnMetLysAsn. 708
4652 CATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCA 4701
709 PheIleAspSerLeuValAlaSerLeuSerAla..... 719
4702 AAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGG 4751
719 719
4752 GTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCG 4801
720 AlaValSerAsnLeuValLysIleLeuLysAspThrAlaA 733
4802 TCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCA 4851
733 laIleAsp.....LeuGluThrArgGlnLysPheGly..... 743
4852 TTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGAC 4901
743 743
4902 GGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTG 4951
743 743
4952 GTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCT 5001
744 ValLeuAspValAlaSerArgLysTrpLeuIleLysPro 756
5002 TGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGG 5051
756 756
5052 GCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCA 5101
756 756
5102 GGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTT 5151
756 756
5152 CCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAA 5201
757 ThrAlaLy 759
5202 ATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGA 5251
759 sSerHisAlaTrpGlyValValGluThrHisAlaArgLys..... 772
5252 GGTCGACATACTCGATA.....TGGTGCGAACAAGATTACGTT 5289
773 TyrHisValAlaLeuLeuGluTyrAspGluGlnGlyValVal 786
5290 AGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACC 5339

```
 787 ThrCysAspAspTrpArgArg..............................  793

5340 TGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAG 5389
                                          :::::::::|||:::
 794 ................................ValAlaValSerSerG  799

5390 ATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACA 5439
     :::::||||||            |||      ::::::      ||||||
 799 luSerValVal............TyrSerAspMetAlaLysLeuArgThr  811

5440 TACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGG 5489
        |||   ||||||          :::      |||::::::|||:::
 812 LeuArgArgLeuLeuArgAsnGlyGluProHisValSerSerAlaLysVa  828

5490 TGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACG 5539
      ||||||
 828 lValLeu.....................................------  830

5540 TTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTG 5589

 830 ------------------------------------------------  830

5590 CGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTAC 5639

 830 ------------------------------------------------  830

5640 CAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCG 5689
                |||||||:::|||
 831 .............ValAspGlyVal.........................  834

5690 TATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCA 5739

 834 ------------------------------------------------  834

5740 CTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGA 5789

 834 ------------------------------------------------  834

5790 ATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAG 5839
                           ||||||  ||||||
 835 ......................ProGlyCysGlyLys.............  839

5840 TGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAAT 5889
                                    |||      |||   ::::::
 840 ............................ThrLysGluIleLeuSerArg  846

5890 GTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAG........... 5928
     |||:::   :::||||||:::   ::::::      |||
 847 ValAsnPheAspGluAspLeuIleLeuValProGlyLysGlnAlaAlaGl  863

5929 ..........AGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACC.... 5964
               |||      |||   :::|||:::      ||||||
 863 uMetIleArgArgArgAlaAsnSerSerGlyIleIleValAlaThrLysA  880

5965 ............ACAGTTAACTCCAGGGTGGTAAACTTT......ATCGTC 5997
                 ||||||:::|||   ::::::||||||      :::
 880 spAsnValLysThrValAspSerPheMetMetAsnPheGlyLysSerThr  896

5998 AGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCA 6047
     |||   ::::::||||||::::::::::||||||      |||:::|||
 897 ArgCysGlnPheLysArgLeuPheIleAspGluGlyLeuMetLeuHisTh  913

6048 AGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCT 6097
      |||   :::::::::   |||      :::   ::::::   :::
 913 rGlyCysValAsnPheLeuValAlaMetSerLeuCysGluIleAlaTyrV  930

6098 TTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTT 6147
       :::||||||   :::||||||||||:::||||||||||   :::   |||
 930 alTyrGlyAspThrGlnGlnIleProTyrIleAsnArgValSerGlyPhe  946

6148 AGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACAC 6197
               :::                :::   :::::::|||
 947 ProTyrProAlaHisPheAlaLysLeuGluValAspGluValGluThrAr  963

6198 TTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA. 6246
           :::   |||||||||   ||||||   :::   |||:::
 963 gArgThrThrLeuArgCysProAlaAspValThrHisTyrLeuAsnArgA  980

6247 ................ATGACCGTAAGGGGAACGGAAAAGTGTTACCCT 6279
                      |||:::::::   :::::::::|||
 980 rgTyrGluGlyPheValMetSerThrSerSerValLysLysSerValSer  996

6280 GAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAA 6329
     ::::::::::|||:::|||             |||:::::::   :::|||||
 997 GlnGluMetValGlyGly......AlaAlaValIleAsnProIleSerLy 1011

6330 AAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGT 6379
     |         |||
1011 sProSerHisGlyLys.............................I 1017

6380 ACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAG 6429
       |||   :::||||||          ||||||   |||:::::|||
1017 leLeuThrPheThrGln.........SerAsp...LysGluAlaLeuLeu 1029

6430 GGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAA 6479
     ::::::|||      :::   |||   ||||||||||||:::|||||::
1030 SerArgGly...TyrSerAspValHisThrValHisGluValGlnGlyGl 1045

6480 AACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCC 6529
     :|||:::|||||||||   |||   |||         :::      |||:
1045 uThrTyrSerAspValSerLeuValArgLeuThrProThrProValSerI 1062

6530 TATTCACT...AAACAACCGCATATACTTGTTGGTTTGTCGAGACACACA 6576
     ::::::::      :::||||||:::||||||:::|||||||||||||||
1062 leIleAlaGlyAspSerProHisValLeuValAlaLeuSerArgHisThr 1078

6577 CGCTCACTGGTTTATGCCGCT..............CTGAGCTCAAAGTT 6611
       ||||||   |||   :::                |||   |||   ::
1079 CysSerLeuLysTyrTyrThrValValMetAspProLeuValSerIleIl 1095

6612 GGACGAT......AAGGTCGGCACATATATTAGCGAC 6642
     :   |||      |||::::::::::|||:::||||||
1095 eArgAspLeuGluLysLeuSerSerTyrLeuSerAsp 1107
```

```
OM of: US-09-301-906-3 to: SPTREMBL_15:*   out_format : pfs

Date: Jan 31, 2001  4:24 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09301906/runat_31012001_153037_9662/app_query.fasta_1.6881
 -DB=SPTREMBL_15 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09301906_@CGN1_1_1354 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-301-906-3
Query length: 6714
Database: SPTREMBL_15:*
Database sequences: 374700
Database length: 117207915
Search time (sec): 906.930000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
sp_virus:O71188       + 3644.00 4224.88  5.4e-228  701  ! O71188 grapevine leafroll-assoc
sp_virus:Q9Q6Q4       +  803.50  923.85  7.0e-44   398  ! Q9q6q4 grapevine leafroll-assoc
sp_virus:O71209       +  802.50  902.35  1.8e-43  2473  ! O71209 grapevine leafroll-assc
sp_virus:Q9Q703       +  746.00  834.08  9.0e-40  3090  ! Q9q703 beet yellows virus. p34
sp_virus:O89906       +  744.50  832.41  1.1e-39  3070  ! O89906 beet yellows virus. fus
sp_virus:Q08534       +  738.50  825.41  2.8e-39  3074  ! Q08534 sugar beet yellow virus
sp_virus:Q66237       +  734.00  819.99  5.4e-39  3124  ! Q66237 citrus tristeza virus.
sp_virus:O40959       +  721.00  808.26  3.3e-38  2301  ! O40959 little cherry closterov
sp_virus:Q9WID7       +  712.50  794.99  1.3e-37  3115  ! Q9wid7 citrus tristeza virus.
sp_virus:Q9IFX0       +  697.50  777.52  1.3e-36  3115  ! Q9ifx0 citrus tristeza virus.
sp_virus:O10467       +  688.00  766.46  5.2e-36  3115  ! O10467 citrus tristeza virus.
sp_virus:P87587       +  668.00  743.20  1.0e-34  3107  ! P87587 citrus tristeza virus.
sp_virus:Q83044       +  650.00  727.88  1.2e-33  1873  ! Q83044 lettuce infectious yell
sp_virus:Q9JGK3       +  486.00  540.70  4.6e-23  1334  ! Q9jgk3 soil-borne wheat mosaic
sp_virus:Q9WIE3       +  484.00  538.24  6.2e-23  1350  ! Q9wie3 chinese wheat mosaic vi
sp_virus:Q9WJD7       +  484.00  534.69  7.1e-23  1858  ! Q9wjd7 chinese wheat mosaic vi
sp_virus:Q9QBU8       +  453.50  502.77  5.9e-21  1345  ! Q9qbu8 oat golden stripe virus
sp_virus:Q9IWA9       +  453.50  499.20  6.7e-21  1853  ! Q9iwa9 oat golden stripe virus
sp_virus:O93163       +  450.50  499.92  9.0e-21  1270  ! O93163 beet soil-borne virus.
sp_virus:Q9JGJ7       +  450.50  499.85  9.0e-21  1277  ! Q9jgj7 sorghum chlorotic spot
sp_virus:O56790       +  450.50  496.17  1.0e-20  1778  ! O56790 beet soil-borne virus.
sp_virus:Q06359       +  447.50  495.99  1.4e-20  1320  ! Q06359 soil-borne wheat mosaic
sp_virus:Q89249       +  447.50  492.37  1.6e-20  1828  ! Q89249 soil-borne wheat mosaic
sp_virus:Q9Q9R0       +  437.50  484.45  6.3e-20  1308  ! Q9q9r0 soil-borne rye mosaic v
sp_virus:Q9JAD4       +  437.50  480.80  7.3e-20  1816  ! Q9jad4 soil-borne rye mosaic v
sp_virus:Q9WK10       +  436.50  483.30  7.4e-20  1306  ! Q9wk10 broad bean necrosis vir
sp_virus:Q9YPH5       +  436.50  479.61  8.5e-20  1820  ! Q9yph5 broad bean necrosis vir
sp_virus:Q9Q9Q6       +  427.50  472.81  2.8e-19  1308  ! Q9q9q6 soil-borne rye mosaic v
sp_virus:Q9JAD2       +  427.50  469.15  3.2e-19  1816  ! Q9jad2 soil-borne rye mosaic v
sp_virus:Q9QCE8       +  423.50  468.15  5.1e-19  1308  ! Q9qce8 european wheat mosaic v
sp_virus:Q9IWB1       +  423.50  464.50  5.9e-19  1816  ! Q9iwb1 european wheat mosaic v
sp_virus:Q9Q9Q8       +  422.50  466.99  6.0e-19  1308  ! Q9q9q8 soil-borne rye mosaic v
sp_virus:Q9JAD3       +  422.50  463.33  6.9e-19  1816  ! Q9jad3 soil-borne rye mosaic v
sp_virus:Q9QBU0       +  413.00  455.98  2.5e-18  1302  ! Q9qbu0 potato mop-top virus. p
sp_virus:Q67704       +  394.00  430.84  4.7e-17  1707  ! Q67704 grapevine virus a. puta
sp_virus:Q96806       +  391.50  427.93  6.8e-17  1707  ! Q96806 grapevine trichovirus b
sp_virus:Q65853       +  386.50  431.87  9.9e-17   710  ! Q65853 beet yellow stunt virus.
sp_virus:Q9WKC2       +  373.00  409.33  9.7e-16  1310  ! Q9wkc2 beet virus q. hypotheti
sp_virus:O90699       +  373.00  405.69  1.1e-15  1818  ! O90699 beet virus q. hypotheti
sp_virus:Q9IH80       +  359.00  388.11  9.5e-15  2038  ! Q9ih80 cherry necrotic rusty m
sp_virus:Q84919       +  354.00  388.37  1.6e-14  1180  ! Q84919 pepper ringspot viru
sp_virus:Q84918       +  354.00  384.32  1.8e-14  1698  ! Q84918 pepper ringspot viru
sp_virus:P90434       +  338.50  370.25  1.6e-13  1188  ! P90434 tobacco rattle virus
sp_virus:P89202       +  337.00  369.06  2.0e-13  1130  ! P89202 sunnhemp mosaic viru
sp_virus:O56250       +  336.50  363.89  2.5e-13  1707  ! O56250 tobacco rattle virus

seq_name: sp_virus:O71188

seq_documentation_block:
ID   O71188      PRELIMINARY;      PRT;   701 AA.
AC   O71188;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HYPOTHETICAL 77.0 KDA PROTEIN (FRAGMENT).
OS   grapevine leafroll-associated virus 3.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=55951;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NY1;
RX   MEDLINE=98264508; PubMed=9603346;
RA   Ling K.-S., Zhu H.-Y., Drong R.F., Slightom J.L., McFerson J.R.,
RA   Gonsalves D.;
RT   "Nucleotide sequence of the 3'-terminal two-thirds of the grapevine
RT   leafroll-associated virus-3 genome reveals a typical monopartite
RT   closterovirus.";
RL   J. Gen. Virol. 79:1299-1307(1998).
DR   EMBL; AF037268; AAC40717.1; -.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   701 AA;  77008 MW;  772A764D2CA2A2DC CRC64;


alignment_scores:
            Quality: 3644.00              Length:    701
              Ratio:  5.198                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-301-906-3 x O71188   ..

 Align seg 1/1  to: O71188  from: 1  to: 701

    4609 GATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTAC 4658
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 AspValTyrAsnGluAlaThrIleAlaThrArgPheSerAsnAlaPheTh 17

    4659 CTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGG 4708
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rPheValAspAsnLeuLysGlyArgSerAlaValPhePheSerLysLeuG 34

    4709 GTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCT 4758
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 lyGluGlyTyrThrTyrAsnGlyGlySerHisValSerSerGlyTrpPro 50

    4759 CGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGA 4808
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ArgAlaLeuGluAspIleLeuThrAlaIleLysTyrProSerValPheAs 67

    4809 CCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACG 4858
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 pHisCysLeuValGlnLysTyrLysMetGlyGlyGlyValProPheHisA 84

    4859 CTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAAT 4908
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 laAspAspGluGluCysTyrProSerAspAsnProIleLeuThrValAsn 100

    4909 CTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGT 4958
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 101 LeuValGlyLysAlaAsnPheSerThrLysCysArgLysGlyGlyLysVa 117

4959 CATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTT 5008
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 117 lMetValIleAsnValAlaSerGlyAspTyrPheLeuMetProCysGlyP 134

5009 TTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATC 5058
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 hcGlnArgThrHisLeuHisSerValAsnSerIleAspGluGlyArgIle 150

5059 AGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTT 5108
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 151 SerLeuThrPheArgAlaThrArgArgValPheGlyValGlyArgMetLe 167

5109 GCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACC 5158
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 167 uGlnLeuAlaGlyGlyValSerAspGluLysSerProGlyValProAsnG 184

5159 AGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGG 5208
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 lnGlnProGlnSerGlnGlyAlaThrArgThrIleThrProLysSerGly 200

5209 GGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGAC 5258
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 GlyLysAlaLeuSerGluGlySerGlyArgGluValLysGlyArgSerTh 217

5259 ATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCA 5308
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 rTyrSerIleTrpCysGluGlnAspTyrValArgLysCysGluTrpLeuA 234

5309 GGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACA 5358
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 rgAlaAspAsnProValMetAlaLeuGluProAspTyrThrProMetThr 250

5359 TTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTT 5408
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 251 PheGluValValLysThrGlyThrSerGluAspAlaValValGluTyrLe 267

5409 GAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGG 5458
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 267 uLysTyrLeuAlaIleGlyIleGluArgThrTyrArgAlaLeuLeuMetA 284

5459 CTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAAT 5508
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 284 laArgAsnIleAlaValThrThrAlaGluGlyValLeuLysValProAsn 300

5509 CAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGA 5558
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 GlnValTyrGluSerLeuProGlyPheHisValTyrLysSerGlyThrAs 317

5559 TCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGT 5608
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 pLeuIlePheHisSerThrGlnAspGlyLeuArgValArgAspLeuProT 334

5609 ACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGAC 5658
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 yrValLeuIleAlaGluLysGlyIlePheThrLysGlyLysAspValAsp 350

5659 GCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTGGT 5708
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 AlaValValAlaLeuGlyAspAsnLeuPheValCysAspAspIleLeuVa 367

5709 TTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCG 5758
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 lPheHisAspAlaIleAsnLeuIleGlyAlaLeuLysValAlaArgCysG 384

5759 GCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCT 5808
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 lyMetValGlyGluSerPheLysSerPheGluTyrLysCysTyrAsnAla 400

5809 CCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGTC 5858
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 ProProGlyGlyGlyLysThrThrThrLeuValAspGluPheValLysSe 417

5859 ACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACA 5908
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 417 rProAsnSerThrAlaThrIleThrAlaAsnValGlySerSerGluAspI 434

5909 TAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGT 5958
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 434 leAsnMetAlaValLysLysArgAspProAsnLeuGluGlyLeuAsnSer 450

5959 GCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTA 6008
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 AlaThrThrValAsnSerArgValValAsnPheIleValArgGlyMetTy 467

6009 TAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTAC 6058
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 467 rLysArgValLeuValAspGluValHisMetMetHisGlnGlyLeuLeuG 484

6059 AACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGAC 6108
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 484 lnLeuGlyValPheAlaThrGlyAlaSerGluGlyLeuPhePheGlyAsp 500

6109 ATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTG 6158
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 501 IleAsnGlnIleProPheIleAsnArgGluLysValPheArgMetAspCy 517

6159 TGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGT 6208
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 517 sAlaValPheValProLysLysGluSerValValTyrThrSerLysSerT 534

6209 ACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGG 6258
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 534 yrArgCysProLeuAspValCysTyrLeuLeuSerSerMetThrValArg 550

6259 GGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACC 6308
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 GlyThrGluLysCysTyrProGluLysValValSerGlyLysAspLysPr 567

6309 AGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAG 6358
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 oValValArgSerLeuSerLysArgProIleGlyThrThrAspAspValA 584

6359 CTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCG 6408
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 584 laGluIleAsnAlaAspValTyrLeuCysMetThrGlnLeuGluLysSer 600

6409 GATATGAAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATGAC 6458
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 AspMetLysArgSerLeuLysGlyLysGlyLysGluThrProValMetTh 617

6459 AGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGA 6508
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 617 rValHisGluAlaGlnGlyLysThrPheSerAspValValLeuPheArgT 634

6509 CGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGTT 6558
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 634 hrLysLysAlaAspAspSerLeuPheThrLysGlnProHisIleLeuVal 650

6559 GGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAA 6608
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 651 GlyLeuSerArgHisThrArgSerLeuValTyrAlaAlaLeuSerSerLy 667

6609 GTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCAG 6658
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 667 sLeuAspAspLysValGlyThrTyrIleSerAspAlaSerProGlnSerV 684

6659 TATCCGACGCTTTGCTTCACACGTTCGCCCCGGCTGGTTGCTTTCGAGGT 6708
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 684 alSerAspAlaLeuLeuHisThrPheAlaProAlaGlyCysPheArgGly 700

6709 ATA 6711
     |||
 701 Ile 701
```

```
seq_name: sp_virus:Q9Q6Q4

seq_documentation_block:
ID   Q9Q6Q4      PRELIMINARY;      PRT;   398 AA.
AC   Q9Q6Q4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HELICASE (FRAGMENT).
GN   HEL.
OS   grapevine leafroll-associated virus 1.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=47985;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fazeli C.F., Rezaian M.A.;
RT   "Nucleotide sequence and organization of ten open reading frames of
RT   the grapevine leafroll-associated virus 1 genome and identification of
RT   three subgenomic RNAs.";
RL   J. Gen. Virol. 0:0-0(2000).
DR   EMBL; AF195822; AAF22737.1; -.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   Helicase.
FT   NON_TER       1      1
SQ   SEQUENCE   398 AA;  44426 MW;  A908109D49C22A23 CRC64;


alignment_scores:
            Quality:  803.50              Length:    394
              Ratio:  2.880                Gaps:      2
 Percent Similarity:  70.812   Percent Identity: 44.162

alignment_block:
 US-09-301-906-3 x Q9Q6Q4   ..

 Align seg 1/1  to: Q9Q6Q4  from: 1  to: 398

    5539 GTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTT 5588
         ||||||   ::::::      ::::::   ::::::   :::||||||
       3 ValTyrAspAlaAsnPheLysValValAlaAsnAlaProGluAspGlyTh 19

    5589 GCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTA 5638
         :::   ||||||   ||||||:::::::::::::   |||   |||:
      19 rThrIlePheAspLeuGluTyrValPheLeuValSerThrGlyThrPheV 36

    5639 CCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTC 5688
         ::      ::::::::::::||||||::::::      |||   ::::::
      36 alProArgArgAsnLeuGlnAlaValLeuSerArgGlnAspAlaValLeu 52

    5689 GTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGC 5738
         |||||||||::::::||||||||||||||||||:::||||||   |||::
      53 ValCysAspGluLeuLeuValPheHisAspAlaMetAsnLeuArgGlyCy 69

    5739 ACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCG 5788
         ::::::::::|||:::   :::||||||||||||   :::|||:::   :
      69 sValArgLeuAlaLysArgAlaMetValGlyGluTyrMetLysAspValA 86

    5789 AATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTA 5838
         ::   ::::::   |||:::||||||||||||||||:::||||||   |||
      86 rgIleSerAlaValAsnSerProProGlyGlyGlyAsnThrThrArgLeu 102

    5839 GTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCTAA 5888
         |||||||||:::          :::   :::          :::||||||
     103 ValAspGluTyrPheGlyArgLysLysArgAlaLysIleAlaAlaAlaAs 119

    5889 TGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGA 5938
         |:::|||||      |||||||||   |||::::::   |||:::   :
     119 nThrGlySerValAlaAspIleAsnAlaAlaIleArgAlaArgGluGlyL 136

    5939 ATTTGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAAC 5988
         ::   |||            |||   |||:::||||||   |||:::|||

     136 ysLysGluProAspLeuValAlaLysThrAlaAsnSerTrpValIleAsn 152

    5989 TTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACAT 6038
              |||                  |||:::|||||||||:::||
     153 SerHisProArgProAsnSerHisValGlyLeuIleAspGluValTyrMe 169

    6039 GATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGG 6088
         |:::|||:::|||:::::::||||||   |||   :::   |||::::::|
     169 tLeuHisLysGlyMetPheGlnLeuThrValValSerMetGlyValLysG 186

    6089 AAGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAG 6138
         ||   :::   :::||||||   ||||||||||||||||||||||||||
     186 luValIleAlaTyrGlyAspLysAsnGlnIleProPheIleAsnArgGlu 202

    6139 AAGGTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGT 6188
         |||:::|||               :::            :::|||:::::
     203 LysThrPheValThrProAsnGluAlaValGluPheAlaGluGluGlnIl 219

    6189 TGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGT 6238
         :   ||||||:::   ||||||||||||||   ||||||||||||:::|
     219 eAspTyrThrAspIleSerTyrArgCysProAlaAspValCysTyrValL 236

    6239 TGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTC 6288
         ||||||||||||||            :::|||   ||||||:::   |||
     236 euSerSerMetThrAspAlaArgGlyLysLysMetTyrProAsnGlyVal 252

    6289 GTTAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAAT 6338
               :::   |||         :::||||||::::::|||   |||||
     253 PheProAsnGlyAspValArgProLeuArgSerPheGluLysValProIl 269

    6339 TGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCA 6388
         |:::|||   :::|||:::         :::|||||||||||||||   |
     269 eAlaThrProGluAspAlaLeuLeuTyrGluAlaAspValTyrLeuThrM 286

    6389 TGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGA..... 6432
         ||||||||   ||||||::::::|||:::::::::::::
     286 etThrGlnAsnGluLysAlaGluMetGlnLysAlaValAlaLysMetGlu 302

    6433 ......AAAGGAAAAGAAACACCA...GTGATGACAGTGCATGAAGCACA 6473
                |||||||:::   |||   |||:::|||:::||||||||||
     303 ValValAlaGlyLysLysArgProAspValIleThrThrHisGluAlaGl 319

    6474 GGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATG 6523
         ||||||||||:::::::::|||||||||   |||   ||||||||||||
     319 nGlyLysThrTyrGluAsnValValLeuValArgLeuLysLysAlaAspA 336

    6524 ACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACAC 6573
         ||   :::|||:::::::::|||||||||:::|||:::||||||||||||
     336 spProIlePheSerArgLysProHisIleValValAlaLeuSerArgHis 352

    6574 ACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGT 6623
         |||||||||:::   ||||||:::||||||||||||:::   |||   ::
     353 ThrArgSerMetLysTyrAlaValLeuSerSerLysMetThrAspThrIl 369

    6624 CGGCACATATATTAGCGACGCGTCACCTCAATCAGTATCCGACGCTTTGC 6673
         ::::      |||:::   :::|||      :::||||||||||:::||||
     369 eSerLysLeuIleAspGlyThrSerAlaGlyLysValSerAspValLeuL 386

    6674 TTCACACGTTCGCCCCGGCTGGTTGCTTTCGA 6705
         ||:::   :::               ||||||
     386 euGlnGlnLeuGlnArgAsnAspArgPheArg 396

seq_name: sp_virus:O71209

seq_documentation_block:
ID   O71209      PRELIMINARY;      PRT;  2473 AA.
AC   O71209;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   METHYLTRANSFERASE/HELICASE POLYPROTEIN (FRAGMENT).
OS   grapevine leafroll-associated virus 2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
```

```
OC   Closterovirus.
OX   NCBI_TaxID=64003;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98264507; PubMed=9603345;
RA   Zhu H.Y., Ling K.S., Goszczynski D.E., McFerson J.R., Gonsalves D.;
RT   "Nucleotide sequence and genome organization of grapevine leafroll-
RT   associated virus-2 are similar to beet yellows virus, the
RT   closterovirus type member.";
RL   J. Gen. Virol. 79:1289-1298(1998).
DR   EMBL; AF039204; AAC40855.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; Methyltransferase; Helicase; Polyprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   2473 AA;  277096 MW;  F8DEEBA6A58ED425 CRC64;


alignment_scores:
            Quality:  802.50              Length:   2475
              Ratio:  0.681                Gaps:    115
 Percent Similarity:  47.596   Percent Identity: 22.424

alignment_block:
 US-09-301-906-3 x O71209   ..

 Align seq 1/1  to: O71209  from: 1  to: 2473

     454 CAAGATGAGCAGAAGCCCAAACGGCAAGCGTCACATTGGGCTGTTAAGCC 503
         |||::::::   |||   ||||||:::   :::|||   :::|||:::
     158 GlnSerServValLysThrLysArgGluAspAlaHisArgThrValGluGl 174

     504 AACAGCTGTTGGCGTCCACGTACCACTTCCTAAAAAACAGGAAGCACTGG 553
             |||:::|||      |||      |||:::::::::::::      |
     174 uArgAlaAlaGlyGlySerValGluGlnProArgGlnLysArgIleAspG 191

     554 AGCCA...GCGCAATCAGTCCCACAACAGTCGTTG..............  585
         ||      :::      ||||||:::   ::::::
     191 luLysGlyCysGlyArgValProSerGlyGlyPheSerHisLeuLeuVal 207

     586 ............GAGGAGAAGGCCGCCTTGACGTTTGGCCTTTTCTTCAG 623
                        |||   :::     ::::::   ||||||:::
     208 GlyAsnLeuAsnGluValArgArgLysValAlaAlaGlyLeuLeuArgPh 224

     624 TAAAGGTGGGGGTGATGAGAGCGACGCTGTCATCTTGCGGAAAGGGAAAT 673
          :::   |||||||||   :::
     224 eArgValGlyGlyAspMetAsp........................... 231

     674 TGTTTAACAGGGCCCTTAATGTTCCTATTGATGTAAAGAACACGTTCGTT 723
           |||:::|||::::::::::::               :::   :::|||
     232 ..PheHisArgSerPheSerThrGlnAlaGly...HisArgLeuLeuVal 246

     724 TGGGCTAAA.....................ATCTGGGATGAAGCCTC 749
         |||   :::                     :::::::::   :::::
     247 TrpArgArgSerSerArgSerValCysLeuGluLeuTyrSerProSerLy 263

     750 TCGTAGGAGAGGGTATTTTTACGTCAAA.........GATAGAGCTGTTA 790
         ::::         |||      ::             |||   |||:::
     263 sAsnPheLeuArgTyrAspValLeuProCysSerGlyAspTyrAlaAlaM 280

     791 AATTCTTCCCTATTGTGCGGGGTAGGGCTACGATCGAGGACTTCATCGTG 840
          |||         :::   ||||||      :::
     280 etPheSerPheAlaAlaGlyGlyArgPheProLeu.............. 291

     841 AATACAGCCCCAGGGTGTGATGTTGCCTTGCCGCGCATTGAGTTGTGGAG 890
                             |||   :::   ||||||:::
     292 .....................ValLeuMetThrArgIleArgTyr..... 299

     891 TATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTA 940
                                       :::|||:::|||::::::
     300 ............................ProAsnGlyPheCysTyrLeuA 307

     941 ACAATGAGAGGCTGAGGGGAGAAATTTACAGACGTCGTTGCTTCTCATCT 990
           :::   |||           :::      |||           :::
     307 laHisCysArgTyrAlaCysAlaPheLeuLeuArgGlyPheAspProLys 323

     991 TCCTTTTCGATAGGT.............................TT 1007
           |||:::||||||                              ::
     324 ArgPheAspIleGlyAlaPheProThrAlaAlaLysLeuArgAsnArgMe 340

    1008 CTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTTGCGGGCA 1057
         ::::   :::||||||   |||||||||
     340 tValSerGluLeuGlyGluArgSerLeu..................... 349

    1058 CGAACATACTACACATGCCATCACTCAATGAAGAGCGTACCTTTGGGTGG 1107
                           :::||||||         ::::::
     350 ...................GlyLeuAsnLeuTyrGlyAlaTyrThrSer 359

    1108 AAGGGC.............................GGAGACGTCTA 1124
         :::|||                                 |||:::
     360 ArgGlyValPheHisCysAspTyrAspAlaLysPheIleLysAspLeuAr 376

    1125 TCTCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGT 1174
          |||              :::|||::::::||||||   :::      :
     376 gLeu............MetSerAlaValIleAlaGly...LysAspGlyV 388

    1175 TGGGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAG 1224
         ::         :::   ::::::::::::   |||::::::|||:::
     388 alGluGluValValProSerAspIleThrProAlaMetLysGlnLys... 403

    1225 GTCTATTCATCGGTCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGA 1274
                       ::::::::::::::   :::||||||
     404 ............ThrIleGluAlaValTyrAspArgLeuTyrGlyGlyTh 416

    1275 CCAATCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTT 1324
          :::|||   |||          :::::      |||   |||:::   :
     416 rAspSerLeuLeuLysLeuSerIleGluLysAspLeuIleAspPheLysA 433

    1325 CTCAAAGGGACGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTG 1374
         :::::   :::::::::   :::::::::|||      :::
     433 snAspValGlnSerLeuLysLysAspArgProIleValLysValProPhe 449

    1375 TTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCT 1424
         :::::::::          :::::::::|||   :::::::::|||:::::
     450 TyrMetSerGluAlaThrGlnAsnSerLeuThrArgPheTyrProGlnPh 466

    1425 TTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATG 1474
         ::::::::::||||||   |||   :::|||:::||||||   |||   |
     466 eGluLeuLysPheSerHisSerSerHisSerAspHisProAlaAlaAlaA 483

    1475 CCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTG 1524
         ||   |||         |||      :::   |||   |||||||||
     483 laSerArgLeuLeuGluAsnGluThrLeuValArgLeuCysGlyAsn... 498

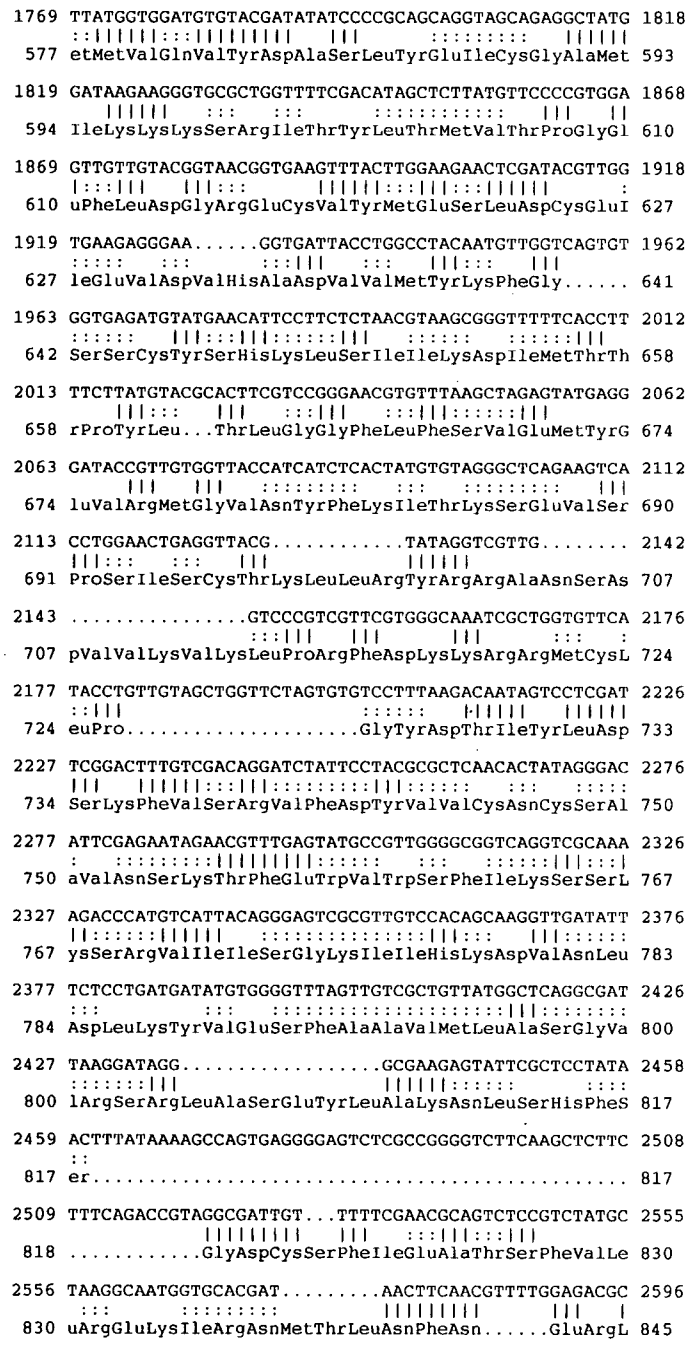
    1525 TGCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCT...GG 1571
                  ||||||||||||         :::|||:::   :::
     499 SerValSerAspIleGlyGlyCysProLeuPheHisLeuHisSerLysTh 515

    1572 CCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGA 1621
          :::   :::   |||||||||:::||||||||||||   ||||||:::
     515 rGlnArgArgValHisValCysArgProValLeuAspGlyLysAspAlaG 532

    1622 AGCGGAGA...ATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGAT 1668
         ::||||||   ::::::::::::   :::   ::::::      ||||||
     532 lnArgArgValValArgAspLeuGlnTyrSerAsnValArgLeuGlyAsp 548

    1669 TCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTC 1718
                  :::|||   :::   |||   :::   ::::::::::
     549 ............AspAspLysIleLeuGlu...GlyProArgAsnIleAs 560

    1719 TTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTT 1768
         :   |||                ||||||   :::::::::|||   :
     560 pIleCysHisTyrProLeuGlyAlaCysAspHisGluSerSerAlaMetM 577
```


1769 TTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATG 1818
577 etMetValGlnValTyrAspAlaSerLeuTyrGluIleCysGlyAlaMet 593
1819 GATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGA 1868
594 IleLysLysLysSerArgIleThrTyrLeuThrMetValThrProGlyGl 610
1869 GTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGG 1918
610 uPheLeuAspGlyArgGluCysValTyrMetGluSerLeuAspCysGluI 627
1919 TGAAGAGGGAA......GGTGATTACCTGGCCTACAATGTTGGTCAGTGT 1962
627 leGluValAspValHisAlaAspValValMetTyrLysPheGly..... 641
1963 GGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTT 2012
642 SerSerCysTyrSerHisLysLeuSerIleIleLysAspIleMetThrTh 658
2013 TTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGG 2062
658 rProTyrLeu...ThrLeuGlyGlyPheLeuPheSerValGluMetTyrG 674
2063 GATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCA 2112
674 luValArgMetGlyValAsnTyrPheLysIleThrLysSerGluValSer 690
2113 CCTGGAACTGAGGTTACG.....TATAGGTCGTTG..... 2142
691 ProSerIleSerCysThrLysLeuLeuArgTyrArgArgAlaAsnSerAs 707
2143 GTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCA 2176
707 pValValLysValLysLeuProArgPheAspLysLysArgArgMetCysL 724
2177 TACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCGAT 2226
724 euPro.....GlyTyrAspThrIleTyrLeuAsp 733
2227 TCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGAC 2276
734 SerLysPheValSerArgValPheAspTyrValValCysAsnCysSerAl 750
2277 ATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAA 2326
750 aValAsnSerLysThrPheGluTrpValTrpSerPheIleLysSerSerL 767
2327 AGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATT 2376
767 ysSerArgValIleIleSerGlyLysIleIleHisLysAspValAsnLeu 783
2377 TCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGAT 2426
784 AspLeuLysTyrValGluSerPheAlaAlaValMetLeuAlaSerGlyVa 800
2427 TAAGGATAGG.....GCGAAGAGTATTCGCTCCTATA 2458
800 lArgSerArgLeuAlaSerGluTyrLeuAlaLysAsnLeuSerHisPheS 817
2459 ACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTC 2508
817 er..... 817
2509 TTTCAGACCGTAGGCGATTGT...TTTTCGAACGCAGTCTCCGTCTATGC 2555
818 GlyAspCysSerPheIleGluAlaThrSerPheValLe 830
2556 TAAGGCAATGGTGCACGAT.....AACTTCAACGTTTTGGAGACGC 2596
830 uArgGluLysIleArgAsnMetThrLeuAsnPheAsn.....GluArgL 845
2597 TTATGTCTATGCCCAGA.....GCGTTCATCCGTAAAGTACCTGGGTCT 2640
845 euLeuGlnLeuValLysArgValAlaPheAlaThrLeuAspValSerPhe 861
2641 GTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAG 2690
862 LeuAspLeuAspSerThrLeuGluSerIleThrAspPheAlaGluCysLy 878
2691 GGGTGCCTTTGATATTTCGAAG.....GAGACCTTCG 2722
878 sValAlaIleGluLeuAspGluLeuGlyCysLeuArgAlaGluAlaGluA 895
2723 GTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTG 2772
895 snGluLysIleArgAsn.....LeuAla 902
2773 GAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGAA.....GA 2813
903 GlyAspSerIleAlaAlaLysLeuAlaSerGluIleValValAspIleAs 919
2814 TGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAATGG 2863
919 pSerLysProSerProLysGln.....ValG 928
2864 GGAACGTTTCTAACGTCCACTGT.....ACGAGGGCAGGT 2898
928 lyAsnSerSerSerGluAsnAlaAspLysArgGluValGlnArgProGly 944
2899 CTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTT 2948
945 LeuArgGlyGlySerArgAsnGlyValValGlyGlu..... 956
2949 GGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCT 2998
957 PheLeuHisPheV 961
2999 TTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGG 3048
961 alValAspSerAlaLeuArgLeuPheLysTyrAlaThrAspGlnGlnArg 977
3049 CTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCAT 3098
978 IleLysSerTyrValArg.....PheLeuAspSerAlaValSerPh 991
3099 ACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAG 3148
991 eLeuAspTyrAsnTyrAspAsnLeuSerPheIleLeuArgValLeuSerG 1008
3149 AATTTTGGAGT..... 3159
1008 luGlyTyrSerCysMetPheAlaPheLeuAlaAsnArgGlyAspLeuSer 1024
3160 GAAGTTTCCTTAAATGA 3176
1025 SerArgValArgSerAlaValCysAlaValLysGluValAlaThrSerCy 1041
3177 CTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGG 3226
1041 sAlaAsnAlaSerValSerLysAlaLysValMetIle..... 1053
3227 CGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGC...GGAATT 3273
1054 ..ThrPheAlaAlaAlaValCysAlaMetMetPheAsnSerCysGlyPhe 1069
3274 GCGGCTAACGCTAGAAGGTACGCGGGC..... 3300
1070 SerGlyAspGlyArgGluTyrLysSerTyrIleHisArgTyrThrGlnVa 1086
3301 AGTAGTTACGAAACTCTAAGCT 3322
1086 lLeuPheAspThrIlePhePheGluAspSerSerTyrLeuProIleGluV 1103
3323 CGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGC... 3369

```
1103 alLeuSerSerAlaIleCysGlyAlaIleValThrLeuPheSerSerGly 1119

3370 .........ACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTC 3410
             ::::::::::::      |||   :::   :::|||||||||
1120 SerSerIleSerLeuAsnAlaPheLeuLeuGlnIleThrLysGlyPheSe 1136

3411 GTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCT 3460
     |      |||:::|||   :::::::::            :::|||
1136 rLeuGluValValValArgAsnValVal............ArgValThrH 1149

3461 TAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTG....... 3503
       :::|||   :::   :::|||:::   :::|||   |||
1149 isGlyLeuSerThrThrAlaThrAspGlyValIleArgGlyValPheSer 1165

3504 CAGTATGTTAGGCATTTG......GGCACATGCTCTTCCACGGCA..... 3542
     |||   |||   ||||||     |||:::   :::::::::|||
1166 GlnIleValSerHisLeuLeuValGlyAsnThrGlyAsnValAlaTyrGl 1182

3543 .....................................CTTA 3546
                                          ::::
1182 nSerAlaPheIleAlaGlyValValProLeuLeuValLysLysCysValS 1199

3547 ATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAA 3596
     ::::::::::   :::|||:::   :::::::::         ::::::
1199 erLeuIlePheIleLeuArgGluAspThrTyrSerGlyPheIleLysHis 1215

3597 TACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAGG 3646
       ::::::   :::   :::|||:::   :::::::::      :::||
1216 GlyIleSerGluPheSerPheLeuSerSerIleLeuLysPheLeuLysGl 1232

3647 TTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCTA 3696
     |   |||:::   |||      ::::::|||                 :
1232 yLysLeuValAspGluLeuLysSerIleIle.................G 1243

3697 GGGGTATCTTTGATCATAAGGGCTTGCTTAACGACACCATACCTCAACT 3746
     ::|||:::||||||   :::      :::   ::::::   |||:::
1243 lnGlyValPheAspSerAsnLysHisValPheLysGluAlaThrGlnGlu 1259

3747 TGCTTACGTCCCACCGGTA.............................. 3765
     ::::::|||   ||| |||
1260 AlaIleArgThrThr.ValMetGlnValProValAlaValValAspAlaL 1276

3766 ...........GAGGGGAGGAATGTGTACGATGAGACGCTTAGGTATTAC 3804
                |||:::         :::   |||   |||
1276 euLysSerAlaAlaGlyLysIleTyrAsnAsnPheThrSerArg...... 1290

3805 CGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGA 3854
     |||   |||      ||||||||:::::   :::|||:::   :::||
1291 ArgThrPheGlyLysAspGluGlySerSerSerAspGlyAlaCysGluGl 1307

3855 AGCGGTTCCCGGTGACGATAACGATGGA...TCCACTTCTAGTGTCTCAA 3901
     |          |||:::::::::|||   :::   ::::::   ||||
1307 uTyrPheSerCysAspGluGlyGluGlyProGlyLeuLysGlyGlySerS 1324

3902 GCTATGATGTTGTCACAAATGTGC............GCGACGTGGGGATT 3939
     |||||    :::|||   :::::                ||||||
1324 erTyrGly.PheSerIleLeuAlaPhePheSerArgIleMetTrpGly.. 1339

3940 AGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAG 3989
           :::   :::|||:::||||||         :::      |||:
1340 .......AlaArgArgLeuIleValLys.........ValLysHisGluC 1351

3990 CGTGCAATACACTTATGTCGAGGAAGAGGTTGCCC............CGT 4027
     ::   :::   |||::::::      |||              |||
1351 ysPheGlyLysLeuPheGluPheLeuSerLeuLysLeuHisGluPheArg 1367

4028 CTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCT 4077
       ::::::::: ||||||:::|||   :::||||||   :::
1368 ThrArgValPhe.GlyLysAsnArgThrAspValGlyValTyrAspPheL 1384

4078 GACGCTATGGCTTTTGTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTT 4127
           ||| :::||||||::::::
1384 euProThrGly.IleValGluThrLeu..................... 1392
```

```
4128 TCACCAACAGTCTAGTGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCA 4177
               ||||||   |||   :::   ::::::|||
1393 ..........SerSerIleGluGluCysAspGlnIleGlu.......... 1402

4178 AAGGGTTGCTCCCAGAAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGG 4227
        ||||||   :::::::::   |||
1403 ..GluLeuLeuGlyAspAspLeuLysGly..................... 1411

4228 GGAAGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGA 4277
                  |||   :::::::::   |||   :::|||      ::
1412 ............AspLysAspAlaSerLeuThrAspMetAsnTyrPheGl 1424

4278 C...AGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATA 4324
     :      |||         :::|||:::   :::   |||   |||   :
1424 uPheSerGluAspPheLeuAlaSerIleGluGluProProPheAlaGlyL 1441

4325 TTCCAAAGGTCACCCAGTCC............................. 4344
     ::         :::::::::
1441 euArgGlyGlySerLysAsnIleAlaIleLeuAlaIleLeuGluTyrAla 1457

4345 ............GAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACC 4382
                 ::::::   |||:::|||      |||:::   :::
1458 HisAsnLeuPheArgIleValAlaSerLysCysSerLysArgProLeuPh 1474

4383 ATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCCC 4432
     ||||||   ::::::   ::::::   ::::::::::::
1474 eLeuAlaPheAlaGluLeuSerSerAlaLeuIleGluLys......PheL 1489

4433 CAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTC 4482
       :::|||      |||                  |||::::::   |||
1489 ysGluValPheProArg.................LysSerGlnLeuVal 1499

4483 GCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAG 4532
     |||   |||   |||      |||   :::      :::   ::::::||
1500 AlaIleValArgGluTyrThrGln...ArgPheLeuArgSerArgMetAr 1515

4533 GAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTA 4582
     |:::|||   :::|||:::      ::: |||:::|||||||||   |||
1515 gAlaLeuGlyLeuAsnAsnGluPheVal.ValLysSerPheAlaAspLeu 1531

4583 GAACGTGTTCCTGCGGTG............TGCAGCTGGATGTGTACAAC 4620
           :::|||||||:::                |||  |||
1532 ......LeuProAlaLeuMetLysArgLysValSer.GlySerPheLeuA 1546

4621 GAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAA 4670
        |||   ||||||      :::            |||:::|||
1546 laSerValTyrArg...ProLeuArgGlyPheSerTyrMetCysValSer 1561

4671 CTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATA 4720
        ||||||:::|||      :::   |||   :::      |||:::
1562 AlaGluArgArgGluLysPhePheAlaLeuValCysLeuIleGlyLeuSe 1578

4721 CCTATAATGGTGGTAGCCATGTTTCATCAGG................... 4751
      |||            |||   |||:::|||
1578 rLeu............ProPhePheValArgIleValGlyAlaLysAlaC 1591

4752 ...........GTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAA 4790
                |||:::|||:::   |||   |||
1591 ysGluGluLeuValSerSerAlaArgArgPheTyr.............. 1602

4791 GTACCCAAGCGTCTTCGA.............................. 4808
           :::|||::::::
1603 ......GluArgIleLysIlePheLeuArgGlnLysTyrValSerLeuSe 1617

4809 .............CCACTGTTTAGTGCAGAAGTACAAGAT.......... 4835
                     |||||||||:::::::|||:::|||
1617 rAsnPhePheCysHisLeuPheSerSerAspValAspAspSerSerAlaS 1634

4835 ......................................................... 4835

1634 erAlaGlyLeuLysGlyGlyAlaSerArgMetThrLeuPheHisLeuLeu 1650
```

```
4836 ..............................GGGTGGAGGCGTACCATTCC 4855
                                   ||||||:::
1651 ValArgLeuAlaSerAlaLeuLeuSerLeuGlyTrpGluGlyLeu..... 1665

4856 ACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATC...TTGACG 4902
               :::   :::|||||||       |||   :::   |||
1666 ..........LysLeuLeuLeuSer.HisHisAsnLeuLeuPheLeuCys 1678

4903 GTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGG 4952
          ||||||      :::|||         |||         |||
1679 PheAlaLeuValAspAspValAsnValLeuIleLysValLeuGlyGlyLe 1695

4953 TAAGGTCATGGTCATAAACGTAGCTTCG..................... 4980
      :::   :::|||      :::   |||
1695 uSerPhePheValGlnProIlePheSerLeuPheAlaAlaMetLeuLeuG 1712

4981 .............GGTGACTAT.......................... 4989
                    :::|||
1712 lnProAspArgPheValGluTyrSerGluLysLeuValThrAlaPheGlu 1728

4990 TTTCTTATGCCTTGC.........................GGTTTTCA 5012
     |||::::::   |||                        ||||||
1729 PhePheLeuLysCysSerProArgAlaProAlaLeuLeuLysGlyPhe.. 1744

5013 AAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTT 5062
                         :::   :::   ::::::   :::|||
1745 ..................PheGluCysValAlaAsnSerThrValSerL 1755

5063 TGACGTTCAGGGCAACTCGGCGCGTCTTT.................GGT 5094
       |||   |||          |||   |||
1755 ysThrValArgArgLeuLeuArgCysPheValLysMetLeuLysLeuArg 1771

5095 GTAGGCAGGATGTTGCAGTTAGCCGGC...GGCGTGTCGGATGAGAAGTC 5141
        ||||||   |||:::      |||   |||:::      :::|||::
1772 LysGlyArgGlyLeuArgAlaAspGlyArgGlyLeuHisArgGlnLysAl 1788

5142 A...CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAA 5188
     :   |||   :::|||:::::::::      ::::::|||:::   |||
1788 aValProValIleProSerAsnArgValValThrAspGlyValGluArgL 1805

5189 CAATCACACCAAAATCGGGGGGCAAGGCTCTATCT.............. 5223
          :::      :::|||   :::||||||
1805 euSerValLysMetGlnGlyValGluAlaLeuArgThrGluLeuArgIle 1821

5224 ..........................GAGGGAAGTGGTAGGGAAGTCAA 5246
                                |||      :::|||:::   ::
1822 LeuGluAspLeuAspSerAlaValIleGluLysLeuAsnArgArg...Ar 1837

5247 GGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAG. 5295
     ::::|||:::|||                 ::::::::::::::|||
1837 gAsnArgAspThrAsn............AspAspGluPheThrArgProA 1850

5296 ..............................TGTGAGTGGCTCAGGGCT 5313
                                    |||:::      :::|||
1850 laHisGluGlnMetGlnGluValThrThrPheCysSer......LysAla 1864

5314 GATAATCCAGTGATGGCTCTTGAA........................ 5337
     ::::::      :::|||||||||
1865 AsnSerAlaGlyLeuAlaLeuGluArgAlaValLeuValGluAspAlaIl 1881

5338 .......CCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGG. 5379
                  :::      :::   |||:::|||:::   |||
1881 eLysSerGluLysLeuSerLysThrValAsnGluMetValArgLysGlyS 1898

5380 .....ACCTCTGAAGATGCCGTCGTG..............GAGTACTTG 5409
          |||||||||:::::::::|||               |||   :::
1898 erThrThrSerGluGluValAlaValAlaLeuSerAspAspGluAlaVal 1914

5410 AAGTATCTGGCTATAGGCATTGAGAGG..............ACATACAG 5444
     :::   ::::::::::::   ||||||               |||   ||
1915 GluGluIleSerValAlaAspGluArgAspAspSerProLysThrValAr 1931

5445 G..............................GCGTTGCTTATGGCTAGAA 5464
```

```
     |                                 ::::::   :::   :::
1931 gIleSerGluTyrLeuAsnArgLeuAsnSerSerPheGluPheProLysP 1948

5465 ATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTT 5514
       |||:::|||          :::   :::      :::   |||   |||
1948 roIleValValAspAspAsnLysAspThrGlyGlyLeuThrAsnAlaVal 1964

5515 TATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCAT 5564
       |||            |||:::      ::::::   :::   :::
1965 ArgGlu............PheTyrTyrMetGlnGluLeuAlaLeuPheGl 1977

5565 TTTTCATTCAACA............CAAGACGGCTTGCGTGTG....... 5595
      :::||||||                  |||   ||||||:::
1977 uIleHisSerLysLeuCysThrTyrTyrAspGlnLeuArgIleValAsnP 1994

5596 ........................AGAGACCTACCGTACGTACTCATA 5619
                              :::|||          :::
1994 heAspArgSerValAlaProCysSerGluAspAlaGlnLeuTyrValArg 2010

5620 GCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCG...GTGGT 5666
       :::   :::   :::::::::||||||:::|||         :::
2011 LysAsnGlySerThrIleValGlnGlyLysGluValArgLeuHisIleLy 2027

5667 AGCTTTGGGCGACAATCTGTTCGTATGCGACGATATACTG......... 5706
       :::   |||:::   |||:::   |||      :::
2027 sAspPheHisAspHisAspPheLeuPheAspGlyLysIleSerIleAsnL 2044

5707 ......................GTTTTCCACGATGCCATTAATTTGATA 5733
                            ::::::||||||      ||||||
2044 ysArgArgArgGlyGlyAsnValLeuTyrHisAsp......AsnLeu... 2057

5734 GGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTT..... 5778
       |||:::   :::::::::      ::::::|||      |||
2058 ...AlaPheLeuAlaSerAsnLeuPheLeuAlaGlyTyrProPheSerAr 2073

5779 ..................AAGTCCTTCGAATATAAGTGCTATAATGCTC 5809
                         :::|||   :::         |||:::||||
2073 gSerPheValPheThrAsnSerSerValAspIleLeuLeuTyrGluAlaP 2090

5810 CCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTT...... 5853
     ||||||||||||||||||||||||||||:::|||:::|||:::
2090 roProGlyGlyGlyLysThrThrThrLeuIleAspSerPheLeuLysVal 2106

5854 ......AAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAG 5897
           |||:::      ||||||      :::|||||||||   :::||
2107 PheLysLysGlyGluValSerThrMetIleLeuThrAlaAsnLysSerSe 2123

5898 TTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAA. 5946
     |:::   :::|||          |||:::|||:::      |||:::|||
2123 rGlnValGluIleLeuLysLysValGluLysGluValSerAsnIleGluC 2140

5947 ........................GGTCTCAACAGTGCTACCACAGTT 5970
                              :::|||      |||:::
2140 ysGlnLysArgLysAspLysArgSerProLysLysSerIleTyrThrIle 2156

5971 AACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTT 6020
     ::::::   :::::::::          ||||||         ||||||
2157 AspAlaTyrLeuMetHisHis......ArgGlyCysAspAlaAspValLe 2171

6021 G...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCG 6067
     |   :::||||||      |||:::|||   |||   :::      :
2171 uPheIleAspGluCysPheMetValHisAlaGlySerValLeuAlaCysI 2188

6068 TCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAG 6117
     ::      |||   :::   :::   ::::::||||||||   :::|||
2188 leGluPheThrArgCysHisLysValMetIlePheGlyAspSerArgGln 2204

6118 ATACCATTCATAAACAGG.........GAGAAGGTGTTTAGGATGGATTG 6158
     |||   :::|||:::|||         :::|||   :::      |||
2205 IleHisTyrIleGluArgAsnGluLeuAspLysCysLeuTyrGlyAspLe 2221

6159 TGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGT 6208
           ||||||         :::      ||||||   :::   ||||
```

```
    2221 uAspArgPheVal...AspLeuGlnCysArgValTyrGlyAsnIleSerT 2237

    6209 ACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATG..... 6249
         ||||||||||   |||||||||      ||||||:::::
    2237 yrArgCysProTrpAspValCysAlaTrpLeuSerThrValTyrGlyAsn 2253

    6250 .........ACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGT 6290
                  ||||||:::|||   ::::::
    2254 LeuIleAlaThrValLysGlyGluSerGlu..................... 2263

    6291 TAGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTG 6340
             ||||||:::               |||:::      :::   |||:
    2264 ....GlyLysSer..............SerMetArgIleAsnGluIleA 2274

    6341 GAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATG 6390
         ::::::::||||||::::::         ::::::::::::||||||||
    2274 snSerValAspAspLeuValProAspValGlySerThrPheLeuCysMet 2290

    6391 ACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGA.. 6438
            |||   ||||||   ::::::::::::   :::      ||||||
    2291 LeuGlnSerGluLysLeuGluIleSerLysHisPheIleArgLysGlyLe 2307

    6439 .AAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCA 6487
             :::      |||:::||||||||||||||||||||:::|||::::
    2307 uThrLysLeuAsnValLeuThrValHisGluAlaGlnGlyGluThrTyrA 2324

    6488 GTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACT 6537
         ::   |||   |||   |||   |||      :::|||:::   |||
    2324 laArgValAsnLeuValArgLeuLysPheGlnGluAspGluProPheLys 2340

    6538 AAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGT 6587
         :::      ||||||   |||:::|||||||||||||||   ||||||::
    2341 SerIleArgHisIleThrValAlaLeuSerArgHisThrAspSerLeuTh 2357

    6588 TTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTA 6637
         :|||   :::|||:::::::::      |||   :::         |||:
    2357 rTyrAsnValLeuAlaAlaArgArgGlyAspAlaThrCysAspAlaIleG 2374

    6638 GCGACGCGTCA 6648
         ::   |||:::
    2374 lnLysAlaAla 2377
```

```
seq_name: sp_virus:Q9Q703

seq_documentation_block:
ID   Q9Q703      PRELIMINARY;      PRT;  3090 AA.
AC   Q9Q703;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P348.
OS   Beet yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BYV-4;
RX   MEDLINE=20079557; PubMed=10611288;
RA   Peremyslov V.V., Hagiwara Y., Dolja V.V.;
RT   "HSP70 homolog functions in cell-to-cell movement of a plant virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BYV-4;
RA   Peremyslov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190581; AAF14300.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   3090 AA;  347990 MW;  09C779A57DE874B9 CRC64;


alignment_scores:
            Quality:  748.00              Length:   2197
              Ratio:  0.740                 Gaps:     97
 Percent Similarity:  46.017   Percent Identity: 22.531

alignment_block:
 US-09-301-906-3 x Q9Q703   ..

 Align seq 1/1  to: Q9Q703  from: 1  to: 3090

    1222 GAGGTCTATTCATCGGTC...........GTTTCGAGTATCACCAATAG 1259
         |||:::   ||||||:::             ::::::|||:::   :::||
     596 GluIleThrSerSerLeuLysHisLysAlaIleGluSerValTyrGluAr 612

    1260 ACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGAAAT 1309
         |:::   ::::::   ::::::   |||         ::::::      |
     612 gValSerIleHisLysAspAsnLeuLeuAlaArgSerValGluLysAspL 629

    1310 TGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCCTCA 1359
         ||   |||:::   ::::::      ::::::   :::||||||   |||
     629 euIleAspPheLysAspGluIleLysSerLeuSerLysGluLysArgSer 645

    1360 CATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGA 1409
                      ::::::              ::::::   |||   ::
     646 ValThrValProPheTyrMetGlyGluAlaValGlnSerGlyLeuThrAr 662

    1410 ACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTC 1459
         :   :::|||:::::::::::::::|||:::   ||||||   |||:::|
     662 gAlaTyrProGlnPheAsnLeuSerPheThrHisSerValTyrSerAspH 679

    1460 ACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGG 1509
         |||||   |||   :::   |||          |||      ::::::
     679 isProAlaAlaAlaGlySerArgLeuLeuGluAsnGluThrLeuAlaSer 695

    1510 AGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCACGTATCA 1559
            :::   :::       |||   ||||||||||||          :::||
     696 MetAlaLysSer...SerPheSerAspIleGlyGlyCysProLeuPheHi 711

    1560 TGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACG 1609
         |:::|||   |||   ::::::   |||||||||:::|||:::   |||:
     711 sIleLysArgGlySerThrAspTyrHisValCysArgProIleTyrAspM 728

    1610 TTAAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCT 1659
         ::||||||::::::||||||:::::::::         |||::::::
     728 etLysAspAlaGlnArgArgValSerArg......GluLeuGlnAlaArg 742

    1660 GGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAA 1709
         |||       :::   :::|||   :::   |||:::||||||   :::::
     743 GlyLeuValGluAsnLeuSerArgGluGlnLeuValGluAlaGlnAlaAr 759

    1710 GTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCG 1759
         :   ||||||   |||   :::         ||||||:::   ::::::|
     759 g...ValSerValCysProHisThrLeuGlyAsnCysAsnValLysSerA 775

    1760 ATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCA 1809
         ||:::   :::||||||:::||||||||   |||   ::::::::::|||
     775 spValLeuIleMetValGlnValTyrAspAlaSerLeuAsnGluIleAla 791

    1810 GAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTT 1859
         :::||||||      |||   :::   |||       ::::::::::::
     792 SerAlaMetValLeuLysGluSerLysValAlaTyrLeuThrMetValTh 808

    1860 CCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTC. 1908
          |||   |||||||||      :::           ::::::   |||
     808 rProGlyGluLeuLeuAspGluArgGluAlaPheAlaIleAspAlaLeuG 825

    1909 .....GATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTT 1953
              |||:::::::|||      :::   |||   :::   |||:::
     825 lyCysAspValValValAspThrArgArgAspMetValGlnTyrLysPhe 841
```

```
1954 GGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTT 2003
     |||      ::::::   |||   |||::::::||||||:::::::::::
 842 Gly......SerSerCysTyrCysHisLysLeuSerAsnIleLysAsnIl 856

2004 T..............TTCACCTTTTCTTATGTACGCACTTCGTCCGGGA 2038
     :             |||||||||                  |||||||
 856 eMetLeuThrProAlaPheThrPhe.................SerGlyA 867

2039 ACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTC 2088
     ||:::|||::::::|||            |||   |||   ::::::
 867 snLeuPheSerValGluMetTyrGluAsnArgMetGlyValAsnTyrTyr 883

2089 ACTATGTGTAGGGCTCAGAAGTCACCTGGA..................... 2118
        :::   |||:::      ||||||
 884 LysIleThrArgSerAlaTyrSerProGluIleArgGlyValLysThrLe 900

2119 ..................ACTGAGGTTACGTATAGGTCGTTGGTCCCGT 2149
                         |||||||||:::      :::|||   |||
 900 uArgTyrArgArgAlaCysThrGluValValGlnValLysLeu...ProA 916

2150 CGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGT 2199
       |||               :::|||:::
 916 rgPheAspLysThrLeuLysThrPheLeuSer..................  926

2200 GTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTA 2249
        ::::::     |||   ||||||:::   ||||||:::|||:::::
 927 ...GlyTyrAspTyrIleTyrLeuAspAlaLysPheValSerArgValPh 942

2250 TTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGT 2299
     ::::|||:::::::::::   ::::::   :::::::::|||||||||:
 942 eAspTyrValValSerAsnCysSerValValAsnSerLysThrPheGluT 959

2300 ATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGT 2349
     :::::   :::   :::::::|||:::|||:::::::|||:::   ::::::
 959 rpValTrpSerTyrIleLysSerSerLysSerArgValValIleSerGly 975

2350 CGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTT 2399
     :::|||:::|||      |||  . |||:::                 ::
 976 LysValIleHisArgAspValHisIleAspLeuLysHisSerGluCysPh 992

2400 AGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGG.............. 2436
     ::::::::::::::|||   ::::::::::::|||
 992 eAlaAlaValMetLeuAlaValGlyValArgSerArgThrThrThrGluP 1009

2437 .....GCGAAGAGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGG 2481
          ||||||::::::::::   |||:::          :::
1009 heLeuAlaLysAsnLeuAsnTyrTyrThrGlyAspAlaSerCysPheGlu 1025

2482 AGTCTCGCCGGGGTCTTCAAGCTCTTC............TTTCAGACCGT 2519
     ::::::      :::|||:::   :::                :::     ::
1026 ThrIleArgPheLeuPheArgGluTrpSerArgArgAlaTyrAlaGluIl 1042

2520 AGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGC 2569
     ::::      |||   :::   ::::::         :::|||   :::
1042 eAsnArgSerPheArgLysLeuMetLysSerIleLeuSerAlaGlyLeuA 1059

2570 ACGATAACTTC............AACGTTTTGGAGACGCTTATGTCTATG 2607
          :::|||             |||   |||:::   |||::::::
1059 spTyrGluPheLeuAspLeuAspAsnSerLeuGlnHisLeuLeuGluTyr 1075

2608 CCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTG 2657
        ::::::   :::|||                    |||:::|||::
1076 SerGluValGluValArg.....................ValSerIleAl 1085

2658 CACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTT 2707
     :   :::|||      |||
1085 aGlnAsnGlyGluValAsp.............................. 1091

2708 CGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTC 2757
                                 |||::::::   ||||||:::
1092 .........................CysAsnGluGluAsnArgValLeu 1099
```

```
2758 TCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGAC 2807
     ::::::   |||:::|||                          ::
1100 ThrGluIleIleAlaGlu.............................Al 1106

2808 AGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCA 2857
     :   |||   |||                      |||:::       :
1106 aAlaAspArgLys.....................SerIleAlaGlnGlyL 1116

2858 TAATGGGGAACGTTTCTAACGTCCACTGT...ACGAGGGCAGGTCTTCTT 2904
     ::   |||   :::|||:::|||            |||:::||||||
1116 euSerGlyAlaLeuSerSerValProThrGlnProArgGlyGlyLeuArg 1132

2905 GGCGGTTCGAAAGCGACC.....................GTGGTTTCGAG 2933
     |||||||||:::   :::                     :::|||:::::
1133 GlyGlySerArgArgSerGlyValSerPheLeuTyrAsnLeuValGluGl 1149

2934 TGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTG 2983
     :|||::::::   :::   :::   |||   |||::::::   :::
1149 uValGlyAsnLeuPhePheSerValGlyAspAlaValArgPheLeuValL 1166

2984 GCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTA 3033
        :::   :::   ||||||   :::   :::|||            :::
1166 ysValPheLysThrPheSerAspSerProIlePheArgValValArgMet 1182

3034 ACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAA 3083
           |||                |||:::::::::   :::   :::
1183 PheLeuAsp..............LeuAlaGluAlaAlaSerProPheVa 1194

3084 CTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGC....... 3126
      |||   |||::::::            |||   |||||||||
1194 lSerValValSerLeuCysAlaTrpLeuArgGluAlaValSerAlaPheS 1211

3127 .................AAGGTCGTAGCTGGAACGAAAGAATTTTGGAGT 3159
                     :::::::|||:::   :::   :::   :::
1211 erSerTrpValAlaAspArgThrValSerGluSerValLysThrPheVal 1227

3160 GAAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCT 3209
     :::   :::::::::   |||   :::|||:::      :::   :::
1228 AsnArgThrValLysArgPheLeuAsnPheMetSerAlaLysThrLeuTh 1244

3210 TATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCA....... 3252
           :::|||         |||:::|||:::   :::|||
1244 rLysLysPhePheArgPhePheLeuSerAlaSerAlaLeuAlaLysThrV 1261

3253 ................................TGGAAG...TATAGG 3264
                                     |||:::   :::
1261 alValArgLysAlaLysValIleLeuGluAlaTyrTrpGluValTrpPhe 1277

3265 CGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAAC 3314
     :::::::|||   :::::::::   :::|||      |||:::   |||
1278 GluSerIleLeuSerAspSerGlyGluTyr......SerAlaValGluPh 1292

3315 TCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCT 3364
        ||||||:::                                :::|||
1292 eCysSerSerValVal.........................IleThrL 1300

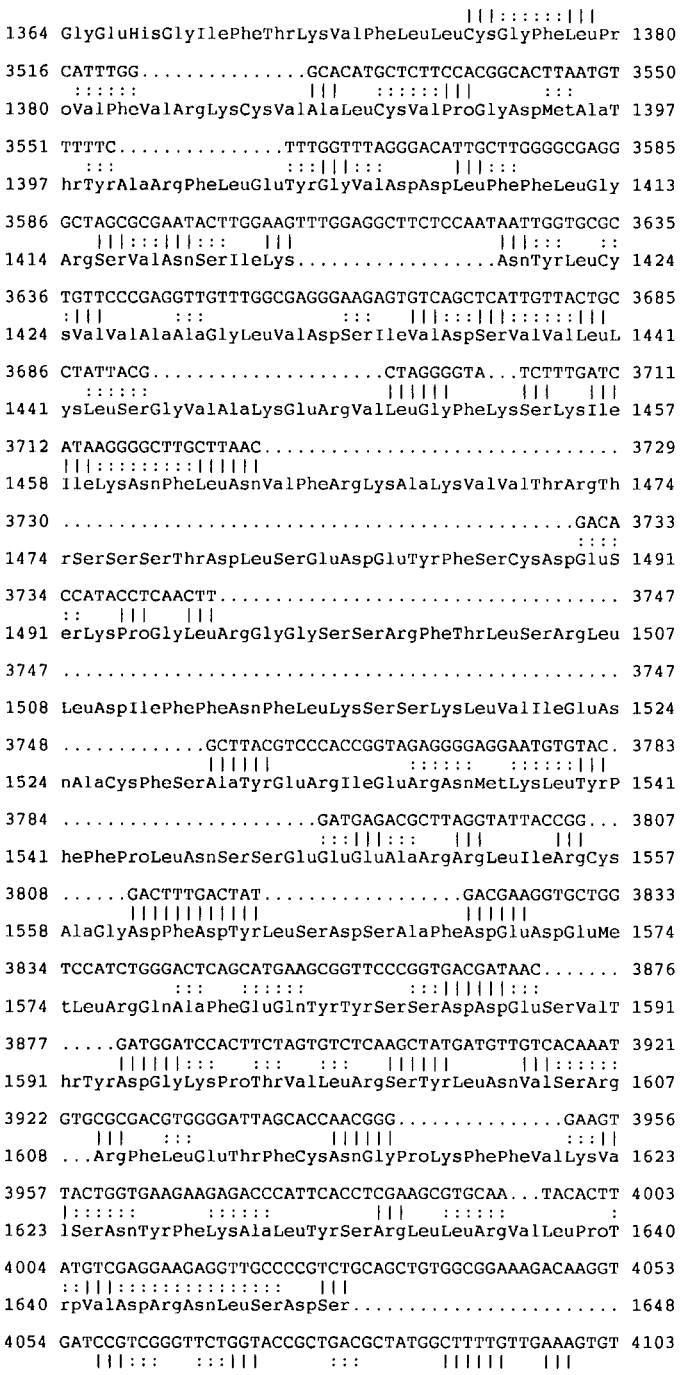
3365 CTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCG 3414
          |||   ||||||      |||            |||||||||
1300 euLeuThrAsnSerGlyArgLeuLeu.........ProGlyPheSerPro 1313

3415 GCGGTGACCGTCACTAGG..............GCGACCGTAGCTAAACG 3449
     ::::::::::::|||:::              ||||||      :::
1314 SerAlaValIleThrGluValLeuLeuAspLeuAlaThrLysIleSerIl 1330

3450 TCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCG 3499
      :::|||   |||      :::|||          |||   |||   |||:
1330 eGluValLeuLeuLysGlnIleSerProValAspSerThrAlaSerSerA 1347

3500 GC.................................................. 3501
     ::
1347 laLeuTyrArgArgValLeuSerGluIleLeuSerAsnPheArgThrMet 1363

3502 ................................TGCAGTATGTTAGG 3515
```


1364 GlyGluHisGlyIlePheThrLysValPheLeuLeuCysGlyPheLeuPr 1380
1380 oValPheValArgLysCysValAlaLeuCysValProGlyAspMetAlaT 1397
1397 hrTyrAlaArgPheLeuGluTyrGlyValAspAspLeuPhePheLeuGly 1413
1414 ArgSerValAsnSerIleLys.....AsnTyrLeuCy 1424
1424 sValValAlaAlaGlyLeuValAspSerIleValAspSerValValLeuL 1441
1441 ysLeuSerGlyValAlaLysGluArgValLeuGlyPheLysSerLysIle 1457
1458 IleLysAsnPheLeuAsnValPheArgLysAlaLysValValThrArgTh 1474
1474 rSerSerSerThrAspLeuSerGluAspGluTyrPheSerCysAspGluS 1491
1491 erLysProGlyLeuArgGlyGlySerSerArgPheThrLeuSerArgLeu 1507
1508 LeuAspIlePhePheAsnPheLeuLysSerSerLysLeuValIleGluAs 1524
1524 nAlaCysPheSerAlaTyrGluArgIleGluArgAsnMetLysLeuTyrP 1541
1541 hePheProLeuAsnSerSerGluGluGluAlaArgArgLeuIleArgCys 1557
1558 AlaGlyAspPheAspTyrLeuSerAspSerAlaPheAspGluAspGluMe 1574
1574 tLeuArgGlnAlaPheGluGlnTyrTyrSerSerAspAspGluSerValT 1591
1591 hrTyrAspGlyLysProThrValLeuArgSerTyrLeuAsnValSerArg 1607
1608 ...ArgPheLeuGluThrPheCysAsnGlyProLysPhePheValLysVa 1623
1623 lSerAsnTyrPheLysAlaLeuTyrSerArgLeuLeuArgValLeuProT 1640
1640 rpValAspArgAsnLeuSerAspSer..... 1648

1649 ...ProGlyLeuLysGlyGlyAsnGluLysAla.LeuLeuAlaLysPheL 1664
1664 euLysThrCysValIleThrAlaCysGluCysValSerGlnIleCysCys 1680
1681 LeuArgLeuIleArgLeuCysTrpGlyThrProAlaCys..... 1693
1694 .GlyLeuValArg.....LeuPheTyrIleT 1702
1702 hrTyrSerGlyThrArgValLeuSerArgValValValAlaValAlaVal 1718
1719 CysProLeuLeu...ValArgAsnGluLeuAspGlyLeuSerAspGlyLe 1734
1734 uThrAsnMetGlyValSerValPheArg..... 1743
1744 ArgLeuPheValAla 1748
1749 LeuArgArgAla..... 1752
1753 LeuS 1754
1754 erAlaTyrSerAsnSerAlaLeuArgArgLysIlePheGluPheIlePhe 1770
1771 GlyAsnIleHisHisProPheAspValAlaValIleGlu..... 1783
1784 ThrAsnGluValAlaProGluProLeuS 1793
1793 erProGluValAspIleAspValAspCysAspPheGlySerAspSerGlu 1809
1810 SerValSerSerAspGluValAlaSerIleProArgProGlyLeuHisGl 1826
1826 yGlySerArgArgSerSerAsnPheLeuThrSerLeuValLysValValP 1843
1843 heLysLeuAlaGlyArgIleProArgLeuLeuPheArgLeuArgAsnPhe 1859
1860 ValAlaTyrPheValGluArgArgLeuAlaSerLysArgLeuLysThrPh 1876
1876 eIleGlyLeuAlaArgLeuPheAspAsnPheSerLeuThrSerVal.... 1891

```
4821 GCAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGG 4870
1892 ........ValTyrLeuLeuGlnGluTyrAspSerValLeuAsnAlaPhe 1905

4871 AGTGCTATCCATCAGATAACCCTATCTTGACGGTCA.............. 4906
1906 IleAspValGluLeuValLeuLeuAsnSerGlySerValAsnValLeuPr 1922

4907 ....ATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGG 4952
1922 oLeuValSerTrpValArgGlySerLeuThrLysLeuAlaGluValIleV 1939

4953 TAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTT 5002
1939 alGlySerGly.....................................Phe 1943

5003 GCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACT.............. 5038
1944 AlaSerPheLeuGlyArgMetCysCysArgValSerAspTrpCysSerSe 1960

5038 .......................................... 5038

1960 rSerSerAsnAlaGlyCysAsnPheMetSerProValArgThrLysGlyL 1977

5039 ..........CCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAA 5077
1977 ysPheValProProSerSerSerGlySerThrAlaSerMetTyrGluArg 1993

5078 CTCGGCGCG......TCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGC 5121
1994 LeuGluAlaLeuGluSerAspIleArgGluHisValLeuSer........ 2007

5122 GGCGTGTCGGATGAGAAGTCACCAG..................... 5146
2008 .ThrCysArgValGlySerAspGluGluGluGluArgProLysGluValT 2024

5147 ..........GTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCA 5185
2024 hrGluProGlyIleGluHisThrSerGlu.........AspValValPro 2037

5186 GAACAATCACAC.................................. 5197
2038 IleArgSerHisSerGlnProLeuSerGlyGlyGluCysSerTyrSerGl 2054

5198 ..........CAAAATCGGGGGGCAAGGCTC......TATCTGAGGGAAG 5231
2054 uAspArgGluGluAsnGluArgAlaAsnLeuLeuProHisValSerLysI 2071

5232 TGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAG 5281
2071 leValSerGluArgArgGlyLeuGluThrAlaArg......ArgAsnLys 2085

5282 ATTACGTTAGGAAGTGTGAGTGG....CTCAGGGCTGATAAT........ 5319
2086 ArgThrLeuArgGlyValSerGluPheLeuAsnAlaIleAsnThrSerAs 2102

5320 ............CCAGTGATGGCTCTTGAACCTGACTACACCCCAATGA 5356
2102 nGluGlnProArgProIleIleVal.........AspHisSerProGluS 2116

5357 CATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTAC 5406
2116 er............ArgAlaLeuThrAsnSerValArgGluPheTyrTyr 2128

5407 TTGAAGTATCTGGCTATA..............GGCATTGAGAGGACATA 5441
2129 LeuGlnGluLeuAlaLeuPheGluLeuSerCysLysLeuArgGluTyrTy 2145

5442 CAGGGCGTTGCTTATGGCTAGA.......................... 5463
2145 rAspGlnLeuLysValAlaAsnPheAsnArgGlnGluCysLeuCysAspL 2162

5464 ..........AATATTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTA 5502
2162 ysAspGluAspMetPheValLeuArgAlaGlyGlnGlyValValSerGly 2178

5503 CCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTACAAGTCGGG 5552
2179 ArgAsn.........SerArgLeuProLeuLysHisPhe......LysGl 2190

5553 CACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACC 5602
2190 yHisGluPheCysPheArgSer......GlyGlyLeu............V 2201

5603 TACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGGGCAAAGAT 5652
2201 alProTyr..................................... 2203

5653 GTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGACGATAT 5702
2204 .......................AspGlyThrSerArgValAspThrIl 2212

5703 ACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTC 5752
2212 e......PheHisThrGlnThrAsnPheValSerAla...AsnAlaLeuL 2226

5753 GATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATAAG...... 5796
2226 euSerGlyTyrLeu......SerTyrArgThrPheThrPheThrAsnLeu 2240

5797 ..............TGCTATAATGCTCCCCCAGGTGGCGGTAAGACGAC 5831
2241 SerAlaAsnValLeuLeuTyrGluAlaProProGlyGlyGlyLysThrTh 2257

5832 GACGTTAGTGGACGAATTCGTTAAG......TCACCCAATAGCACAGCCA 5875
2257 rThrLeuIleLysValPheCysGluThrPheSerLysValAsnSerLeuI 2274

5876 CCATTACGGCTAATGTGGGAAGTTCTGAGGACATA.............. 5910
2274 leLeuThrAlaAsnLysSerSerArgGluGluIleLeuAlaLysValAsn 2290

5911 AATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGC 5960
2291 ArgIleValLeuAspGluGlyAspThrProLeuGlnThrArgAspArgIl 2307

5961 TACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATA 6010
2307 eLeuThrIleAspSerTyrLeuMetAsn......AsnArgGlyLeuThrC 2322

6011 AAAGGGTTTTG...GTGGATGAGGTGCACATGATGCATCAAGGCTTACTA 6057
2322 ysLysValLeuTyrLeuAspGluCysPheMetValHisAlaGlyAlaAla 2338

6058 CAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGA 6107
2339 ValAlaCysIleGluPheThrLysCysAspSerAlaIleLeuPheGlyAs 2355

6108 CATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATT 6157
2355 pSerArgGlnIleHisTyrIleAspArgAsnGlu.........LeuAspT 2369

6158 GTGCTGTT.................TTTGTTCCAAAGAAGGAAAGCGTT 6189
2369 hrAlaValLeuSerAspLeuAsnArgPheVal...AspAspGluSerArg 2384

6190 GTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTT 6239
2385 ValTyrGlyGluValSerTyrArgCysProTrpAspValCysAlaTrpLe 2401

6240 GTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCG 6289
2401 uSerThrPhe.....................TyrProLysThrValA 2410

6290 TT.................AGCGGTAAGGACAAACCAGTAGTAAGATCG 6321
```

```
      ::                  :::|||::::::              |||
  2410 laThrThrAsnLeuValSerAlaGlyGlnSer.............Ser 2421

  6322 CTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGC 6371
       ::::::   |||   |||   ::::::|||||||||        :::
  2422 MetGlnValArgGluIleGluSerValAspAspValGluTyrSerSerGl 2438

  6372 TGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATATG..... 6414
          |||||||||   |||   |||   ||||||:::|||:::
  2438 uPheValTyrLeuThrMetLeuGlnSerGluLysLysAspLeuLeuLysS 2455

  6415 ........AAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACACCAGTGATG 6456
               |||||||||   ::::::      |||   |||   |||:::
  2455 crPheGlyLysArgSerArgSerSerValGluLysProThr...ValLeu 2470

  6457 ACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAG 6506
       |||||||||||||||||||||:::|||:::      |||   |||   ||
  2471 ThrValHisGluAlaGlnGlyGluThrTyrArgLysValAsnLeuValAr 2487

  6507 GACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTG 6556
       |||||||      :::|||:::   |||   ::::::   ||||||   |
  2487 gThrLysPheGlnGluAspAspProPheArgSerGluAsnHisIleThrV 2504

  6557 TTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCA 6606
       ||:::|||||||||||||::::::||||||:::|||::::::|||||||||
  2504 alAlaLeuSerArgHisValGluSerLeuThrTyrSerValLeuSerSer 2520

  6607 AAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCG 6645
       |||   ||||||   ::::::      |||      |||
  2521 LysArgAspAspAlaIleAlaGlnAlaIleValLysAla 2533

seq_name: sp_virus:O89906

seq_documentation_block:
ID   O89906      PRELIMINARY;      PRT;  3070 AA.
AC   O89906;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   FUSION PROTEIN.
OS   Beet yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CALIFORNIAN;
RX   MEDLINE=98285746; PubMed=9621048;
RA   Peremyslov V.V., Hagiwara Y., Dolja V.V.;
RT   "Genes required for replication of the 15.5-kilobase RNA genome of a
RT   plant closterovirus.";
RL   J. Virol. 72:5870-5876(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CALIFORNIAN;
RA   Dolja V., Peremyslov V., Hagivara Y.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF056575; AAC25115.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
FT   CHAIN         1    588       PAPAIN-LIKE LEADER PEPTIDE.
FT   CHAIN       589   2042       PUTATIVE METHYLTRANSFERASE AND RNA
FT                                HELICASE.
FT   CHAIN      2626   3070       RNA-DEPENDENT RNA POLYMERASE.
SQ   SEQUENCE   3070 AA;  345803 MW;  6FC307093F1DF36E CRC64;


alignment_scores:
            Quality:  744.50             Length:   2220
                Ratio:  0.731              Gaps:    103
 Percent Similarity:  45.856   Percent Identity: 22.252

alignment_block:
 US-09-301-906-3 x O89906   ..

 Align seg 1/1  to: O89906  from: 1  to: 3070

   1222 GAGGTCTATTCATCGGTC.....GTTTCGAGTATCACCAATAG 1259
        |||:::   ||||||:::           ::::::|||:::   :::||
    596 GluIleThrSerSerLeuLysHisLysAlaIleGluSerValTyrGluAr 612

   1260 ACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACACGAAAT 1309
        |:::   ::::::   ::::::   |||         ::::::      |
    612 gValSerIleHisLysAspAsnLeuLeuAlaArgSerValGluLysAspL 629

   1310 TGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAACCCTCA 1359
        ||   |||:::   ::::::      ::::::   :::||||||   |||
    629 euIleAspPheLysAspGluIleLysSerLeuSerLysGluLysArgSer 645

   1360 CATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGA 1409
                ::::::            ::::::   |||   ::
    646 ValThrValProPheTyrMetGlyGluAlaValGlnSerGlyLeuThrAr 662

   1410 ACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAGTAGTC 1459
        :   :::|||:::::::::::::::|||:::   ||||||   |||:::|
    662 gAlaTyrProGlnPheAsnLeuSerPheThrHisSerValTyrSerAspH 679

   1460 ACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTTCCAGG 1509
        |||||   |||   :::   |||         |||      ::::::
    679 isProAlaAlaAlaGlySerArgLeuLeuGluAsnGluThrLeuAlaSer 695

   1510 AGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCACGTATCA 1559
           :::   :::      |||   ||||||||||||         :::||
    696 MetAlaLysSer...SerPheSerAspIleGlyGlyCysProLeuPheHi 711

   1560 TGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCTAGACG 1609
        |:::|||   |||   ::::::   |||||||||:::|||:::   |||:
    711 sIleLysArgGlySerThrAspTyrHisValCysArgProIleTyrAspM 728

   1610 TTAAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCT 1659
        ::||||||::::::||||||:::::::::         |||::::::
    728 etLysAspAlaGlnArgArgValSerArg......GluLeuGlnAlaArg 742

   1660 GGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAA 1709
        |||      :::   :::|||   :::   |||:::||||||   :::::
    743 GlyLeuValGluAsnLeuSerArgGluGlnLeuValGluAlaGlnAlaAr 759

   1710 GTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCG 1759
        :   ||||||   |||   ::          ||||||:::   ::::::|
    759 g...ValSerValCysProHisThrLeuGlyAsnCysAsnValLysSerA 775

   1760 ATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCA 1809
        ||:::   :::||||||:::|||||||||   |||   :::::::::|||
    775 spValLeuIleMetValGlnValTyrAspAlaSerLeuAsnGluIleAla 791

   1810 GAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTT 1859
        :::|||||||      |||   :::   |||      ::::::::::::
    792 SerAlaMetValLeuLysGluSerLysValAlaTyrLeuThrMetValTh 808

   1860 CCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTC. 1908
         |||   |||||||||      :::                ::::::   |||
    808 rProGlyGluLeuLeuAspGluArgGluAlaPheAlaIleAspAlaLeuG 825

   1909 .....GATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTT 1953
             |||:::::::|||      :::   |||   :::   |||:::
    825 lyCysAspValValValAspThrArgArgAspMetValGlnTyrLysPhe 841

   1954 GGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTT 2003
        |||      ::::::   |||   |||:::::::||||||::::::::::
    842 Gly......SerSerCysTyrCysHisLysLeuSerAsnIleLysAsnIl 856

   2004 T..............TTCACCTTTTCTTATGTACGCACTTCGTCCGCGA 2038
```

```
    :              |||||||||                    |||||||
856 eMetLeuThrProAlaPheThrPhe.................SerGlyA 867

2039 ACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTC 2088
     ||:::|||:::::|||            |||   |||   ::::::
 867 snLeuPheSerValGluMetTyrGluAsnArgMetGlyValAsnTyrTyr 883

2089 ACTATGTGTAGGGCTCAGAAGTCACCTGGA..................... 2118
       :::   |||:::      ||||||
 884 LysIleThrArgSerAlaTyrSerProGluIleArgGlyValLysThrLe 900

2119 ...................ACTGAGGTTACGTATAGGTCGTTGGTCCCGT 2149
                         |||||||||:::      :::|||   |||
 900 uArgTyrArgArgAlaCysThrGluValValGlnValLysLeu...ProA 916

2150 CGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGT 2199
       |||               :::|||:::
 916 rgPheAspLysThrLeuLysThrPheLeuSer.................. 926

2200 GTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTA 2249
       ::::::      |||   ||||||:::   ||||||:::|||:::::
 927 ...GlyTyrAspTyrIleTyrLeuAspAlaLysPheValSerArgValPh 942

2250 TTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGT 2299
     ::::|||:::::::::::   ::::::   :::::::::|||||||||:
 942 eAspTyrValValSerAsnCysSerValValAsnSerLysThrPheGluT 959

2300 ATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGT 2349
     :::::   :::   ::::::|||:::|||:::::::|||:::   ::::::
 959 rpValTrpSerTyrIleLysSerSerLysSerArgValValIleSerGly 975

2350 CGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTT 2399
     :::|||:::|||      |||   |||:::                  ::
 976 LysValIleHisArgAspValHisIleAspLeuLysHisSerGluCysPh 992

2400 AGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGG.............. 2436
     ::::::::::::::::|||   ::::::::::::|||
 992 eAlaAlaValMetLeuAlaValGlyValArgSerArgThrThrThrGluP 1009

2437 .....GCGAAGAGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGG 2481
          ||||||::::::::::   |||:::          :::
1009 heLeuAlaLysAsnLeuAsnTyrTyrThrGlyAspAlaSerCysPheGlu 1025

2482 AGTCTCGCCGGGGTCTTCAAGCTCTTC............TTTCAGACCGT 2519
     ::::::      :::|||:::   :::                ::      ::
1026 ThrIleArgPheLeuPheArgGluTrpSerArgArgAlaTyrAlaGluIl 1042

2520 AGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGC 2569
     ::::      |||   :::   ::::::          :::||||   :::
1042 eAsnArgSerPheArgLysLeuMetLysSerIleLeuSerAlaGlyLeuA 1059

2570 ACGATAACTTC............AACGTTTTGGAGACGCTTATGTCTATG 2607
          :::|||            |||   |||:::   |||::::::
1059 spTyrGluPheLeuAspLeuAspAsnSerLeuGlnHisLeuLeuGluTyr 1075

2608 CCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTG 2657
       ::::::   :::|||                     |||:::|||::
1076 SerGluValGluValArg.....................ValSerIleAl 1085

2658 CACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTT 2707
     :   :::|||      |||
1085 aGlnAsnGlyGluValAsp.............................. 1091

2708 CGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTC 2757
                                 |||::::::   ||||||:::
1092 .........................CysAsnGluGluAsnArgValLeu 1099

2758 TCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGAC 2807
     ::::::   |||:::|||                              ::
1100 ThrGluIleIleAlaGlu............................Al 1106

2808 AGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCA 2857
     :   |||   |||                     |||:::        :

1106 aAlaAspArgLys.....................SerIleAlaGlnGlyL 1116

2858 TAATGGGGAACGTTTCTAACGTCCACTGT...ACGAGGGCAGGTCTTCTT 2904
     ::   |||   :::|||:::|||            |||:::||||||
1116 euSerGlyAlaLeuSerSerValProThrGlnProArgGlyGlyLeuArg 1132

2905 GGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGC 2954
     |||||||||:::
1133 GlyGlySerArgArg................................... 1137

2955 TCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCC. 3003
                                 ||||||::::::   :::::::::
1138 .........................SerGlyValSerPheLeuTyrAsnL 1146

3004 ...........ACAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGAT 3042
                 :::|||:::||||||::::::   |||
1146 euValGluGluValGlyAsnLeuPhePheSerValGly............ 1158

3043 GAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGT 3092
                                       :::|||::::::     ||
1159 ..............................AspAlaValArgPheLeuVa 1165

3093 GGGCATACTGGAGACG...TCGCGCGTAGCTGTGAGCAAGGTCGTA.... 3135
     |   ::::::::::|||   |||         :::   :::||||||
1165 lLysValPheLysThrPheSerAspSerProIlePheArgValValArgM 1182

3136 .................GCTGGAACGAAAGAATTTTGGAGTGAAGTTTCC 3168
                       :::::::::,  |||   |||   ||||||
1182 etPheLeuAspLeuAlaGluAlaAlaSerProPheValSerValValSer 1198

3169 TTAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGAT 3218
     |||      :::            ||||||:::   |||
1199 LeuCysAlaTrp............LeuArgGluAlaVal............ 1207

3219 ATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCG 3268
                       :::|||      :::|||         |||
1208 ..................SerAlaPheSerSerTrpValAlaAspArgT 1218

3269 GAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTA 3318
       ::::::   :::::::::   :::::::::   :::   :::   :::
1218 hrValSerGluSerValLysThrPheValAsnArgThrValLysArgPhe 1234

3319 AGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAG 3368
       ::::::   |||                  :::   ||||||:::::
1235 LeuAsnPheMetSer..............AlaLysThrLeuThrLysLy 1246

3369 CACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGG 3418
     :                              |||   |||   ||||||
1246 sPhePhe.......................ArgPhePheLeuSerAlaS 1255

3419 TGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCTTA...... 3462
       ::::::::::::::::::|||   |||   :::|||   |||
1255 erAlaLeuAlaLysThrValValArgLysAlaLysValIleLeuGluAla 1271

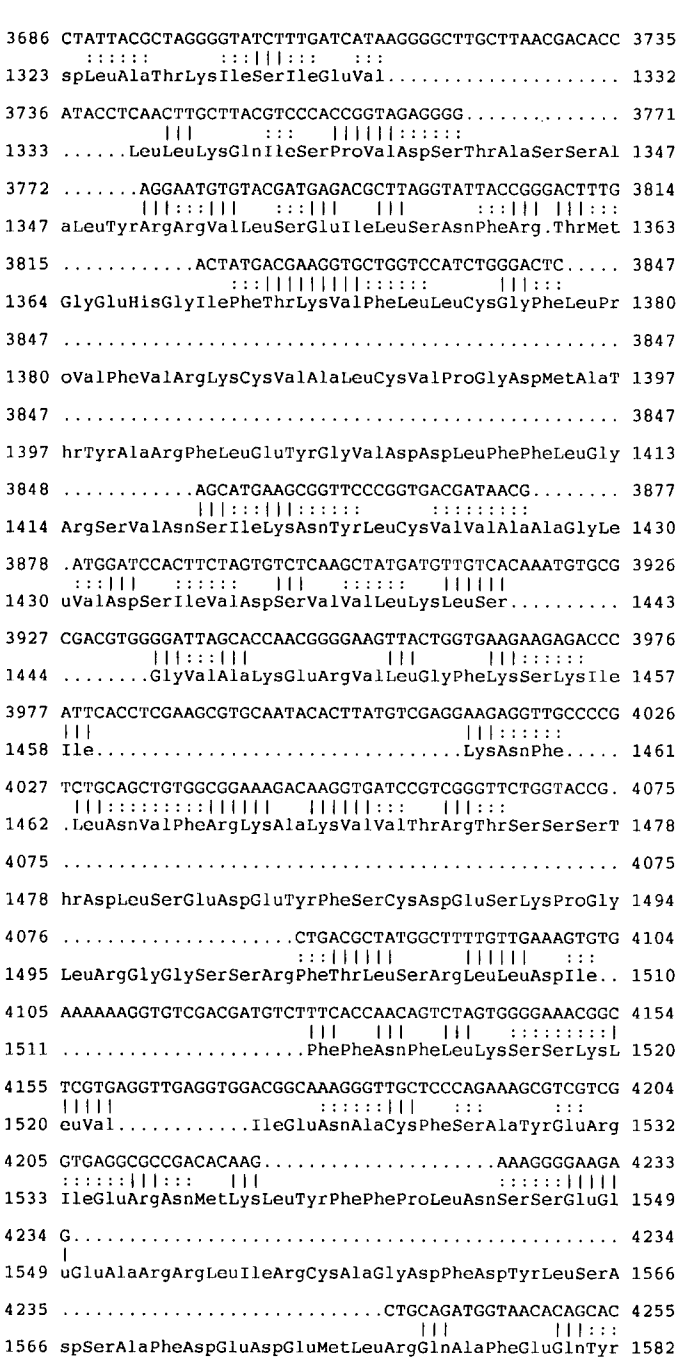
3463 .....................GCGTTGCTATCGTTTTCTACCTCATAC.. 3489
                           ::::::||||||   |||   :::|||
1272 TyrTrpGluValTrpPheGluSerIleLeuSerAspSerGlyGluTyrSe 1288

3490 .GCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCTTCCAC 3538
      |||::::::   ||||||   :::   |||
1288 rAlaValGluPheCysSerSerValValIle................... 1298

3539 GGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCT 3588
                                    |||||||||
1299 ............................ThrLeuLeu............ 1301

3589 AGCGCGAATACTTGGAAG...TTTGGAGGCTTCTCCAATAATTGGTGCGC 3635
       :::|||:::   :::   :::   |||||||||
1302 ...ThrAsnSerGlyArgLeuLeuProGlyPheSer.............. 1312

3636 TGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGC 3685
       |||::::::|||                 :::::::::::::||||||
1313 ....ProSerAlaVal..............IleThrGluValLeuLeuA 1323
```


3686 CTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACC 3735
1323 spLeuAlaThrLysIleSerIleGluVal..... 1332
3736 ATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGG..... 3771
1333 LeuLeuLysGlnIleSerProValAspSerThrAlaSerSerAl 1347
3772 AGGAATGTGTACGATGAGACGCTTAGGTATTACCGGGACTTTG 3814
1347 aLeuTyrArgArgValLeuSerGluIleLeuSerAsnPheArg.ThrMet 1363
3815 ACTATGACGAAGGTGCTGGTCCATCTGGGACTC..... 3847
1364 GlyGluHisGlyIlePheThrLysValPheLeuLeuCysGlyPheLeuPr 1380
3847 3847
1380 oValPheValArgLysCysValAlaLeuCysValProGlyAspMetAlaT 1397
3847 3847
1397 hrTyrAlaArgPheLeuGluTyrGlyValAspAspLeuPhePheLeuGly 1413
3848 AGCATGAAGCGGTTCCCGGTGACGATAACG..... 3877
1414 ArgSerValAsnSerIleLysAsnTyrLeuCysValValAlaAlaGlyLe 1430
3878 .ATGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCG 3926
1430 uValAspSerIleValAspSerValValLeuLysLeuSer..... 1443
3927 CGACGTGGGGATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCC 3976
1444 GlyValAlaLysGluArgValLeuGlyPheLysSerLysIle 1457
3977 ATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCG 4026
1458 Ile.....LysAsnPhe..... 1461
4027 TCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCG. 4075
1462 .LeuAsnValPheArgLysAlaLysValValThrArgThrSerSerSerT 1478
4075 4075
1478 hrAspLeuSerGluAspGluTyrPheSerCysAspGluSerLysProGly 1494
4076 CTGACGCTATGGCTTTTGTTGAAAGTGTG 4104
1495 LeuArgGlyGlySerSerArgPheThrLeuSerArgLeuLeuAspIle.. 1510
4105 AAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGC 4154
1511 PhePheAsnPheLeuLysSerSerLysL 1520
4155 TCGTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCG 4204
1520 euVal.....IleGluAsnAlaCysPheSerAlaTyrGluArg 1532
4205 GTGAGGCGCCGACACAAG.....AAAGGGGAAGA 4233
1533 IleGluArgAsnMetLysLeuTyrPhePheProLeuAsnSerSerGluGl 1549
4234 G..... 4234
1549 uGluAlaArgArgLeuIleArgCysAlaGlyAspPheAspTyrLeuSerA 1566
4235 CTGCAGATGGTAACACAGCAC 4255
1566 spSerAlaPheAspGluAspGluMetLeuArgGlnAlaPheGluGlnTyr 1582
4256 AAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTC..... 4300
1583 TyrSerSerAspAspGluSerValThrTyrAspGlyLysProThrValLe 1599
4301 TTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCA 4340
1599 uArgSerTyrLeuAsnValSerArgArgPheLeuGluThrPheCysAsnG 1616
4341 GTCCGAGGTACATGCTCAGA...AAGAAGTGAAACAAGAAGTACCATTGG 4387
1616 lyProLysPhePheValLysValSerAsnTyrPheLysAlaLeuTyrSer 1632
4388 CGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCCCCAAGT 4437
1633 ArgLeuLeuArg..ValLeuProTrpValAspArgAsnLeuSerAspSer 1648
4438 GTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCA 4487
1649 ProGlyLeuLysGly.....GlyAsnGluLysAlaLe 1659
4488 TGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTT 4537
1659 uLeuAlaLys.....PheLeuLysThrC 1667
4538 TGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGA... 4584
1667 ysValIleThrAlaCysGluCysValSerGlnIleCysCysLeuArgLeu 1683
4585 ACGTGTTCCTGCGGT...GTGCAGCTG..... 4608
1684 IleArgLeuCysTrpGlyThrProAlaCysGlyLeuValArgLeuPheTy 1700
4609 .GATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTA 4657
1700 rIleThrTyrSerGlyThrArgValLeuSerArgValValValAlaValA 1717
4658 CC.....TTTGTCGATAAC 4671
1717 laValCysProLeuLeuValArgAsnGluLeuAspGlyLeuSerAspGly 1733
4672 TTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATAC 4721
1734 LeuThrAsnMetGlyValSerValPheArgArgLeuPheValAlaLeuAr 1750
4722 CTATAATGGTGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTA...G 4768
1750 gArgAlaLeuSerAlaTyrSerAsnSerAlaLeuArgArgLysIlePheG 1767
4769 AGGATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTA 4818
1767 luPheIlePheGlyAsnIleHisHisPro.....PheAspValAlaVal 1781
4819 GTGCAGAAGTACAAGATG..... 4836
1782 IleGluThrAsnGluValAlaProGluProLeuSerProGluValAspIl 1798
4837 GGTGGAGGCGTACCATTCCACGCTGATGACG 4867
1798 eAspValAspCysAspPheGlySerAspSerGluSerValSerSerAspG 1815
4868 AGGAGTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGG 4917
1815 lu..ValAlaSerIleProArgProGlyLeuHisGlyGlySerArg.Arg 1830
4918 AAGGCAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCAT 4967
1831 SerSerAsnPheLeuThrSerLeuValLys.....ValValPh 1843
4968 AAACGTAGCTTCGGGT..... 4983
1843 eLysLeuAlaGlyArgIleProArgLeuLeuPheArgLeuArgAsnPheV 1860
4984 ..GACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCA 5031


1860 alAlaTyrPheVal 1864
5032 GTAAACTCCATCGACGAAGGGCGCATCAGT TTGACGTT 5069
1865 GluArgArgLeuAlaSerLysArgLeuLysThrPh 1876
5070 CAGGGCAACTCGGCGCGTCTTT GGTGTAGGCA 5101
1876 eIleGlyLeuAlaArgLeuPheAspAsnPheSerLeuThrSerValValT 1893
5102 GGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAG 5136
1893 yrLeuLeuGlnGluTyrAspSerValLeuAsnAlaPheIleAspValGlu 1909
5137 AAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGG 5177
1910 LeuValLeuLeuAsnSerGlySerValAsnValLeuProLeuValSerTr 1926
5178 TGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAG 5226
1926 pValArgGlySerLeuThr LysLeuAlaGluV 1937
5227 GGAAGTGGT AGGGAAGTCAAGGGGAGG TCGACA 5259
1937 alIleValGlySerGlyPheAlaSerPheLeuGlyArgMetCysCysArg 1953
5260 TACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAG 5309
1954 ValSerAspTrpCysSerSerSerSerAsnAlaGlyCysAsnPheMet 1969
5310 GGCTGATAATCCAGTGATG 5328
1970 SerProValArgThrLysGlyLysPheValProProSerSerS 1984
5329 GCTCTTGAACCTGAC 5343
1984 erGlySerThrAlaSerMetTyrGluArgLeuGluAlaLeuGluSerAsp 2000
5344 TACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGC 5393
2001 IleArgGluHisValLeuSerThrCysArgValGlySerAspGluGluGl 2017
5394 CGTC GTGGAGTACTTGAAGTATC 5416
2017 uGluArgProLysGluValThrGluProGlyIleGluHisThrSerGluA 2034
5417 TGGCTATAGGCATTGAGAGGACATACA GGGCGTTG 5451
2034 spValValProIle.ArgSerHisSerGlnProLeuSerGlyGlyGluCy 2050
5452 CTTATGGCTAGAA ATATTGCCGT 5474
2050 sSerTyrSerGluAspArgGluGluAsnGluArgAlaAsnLeuLeuPro 2066
5475 CACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCAC 5524
2067 HisValSerLysIleValSerGlu 2074
5525 TACCGGGCTTTCACGTTTACAAGTCGGGCACAG AT 5559
2075 ArgArgGlyLeuGluThrAlaArgArgAsnLysArgThrLeuArgGlyVa 2091
5560 CTCATTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCT 5603
2091 lSerGluPheLeu.AsnAlaIleAsnThrSerAsnGluGlnProArgPro 2107
5604 ACCGTACGTACTCATAGCTGAAA 5626
2108 IleIleValAspHisSerProGluSerArgAlaLeuThrAsnSerValAr 2124
5627 AAGGTATCTTTACCAAGGGCAA AGATGTCGA 5657
2124 gGluPheTyrTyrLeuGlnGluLeuAlaLeuPheGluLeuSerCysLysL 2141
5658 CGCGGTGGTAGCTTTGGGCGACAATCT 5684
2141 euArgGluTyrTyrAspGlnLeuLysValAlaAsnPheAsnArgGlnGlu 2157
5685 GTTCGTATGCGAC GATATACTGGTT 5709
2158 Cys.LeuCysAspLysAspGluAspMetPheValLeuArgAlaGlyGlnG 2174
5709 5709
2174 lyValValSerGlyArgAsnSerArgLeuProLeuLysHisPheLysAsp 2190
5709 5709
2191 HisGluPheCysPheArgSerGlyGlyLeuValProTyrAspGlyThrSe 2207
5710 TTCCACGATGCCATTAATTTGATAGGTGCACTGA 5743
2207 rArgValAspThrIlePheHisThrGlnThrAsnPheValSerAla A 2223
5744 AAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATAT 5793
2223 snAlaLeuLeuSerGlyTyrLeu SerTyrArgThrPheThrPhe 2237
5794 AAG TGCTATAATGCTCCCCCAGGTGGCGG 5822
2238 ThrAsnLeuSerAlaAsnValLeuLeuTyrGluAlaProProGlyGlyGl 2254
5823 TAAGACGACGACGTTAGTGGACGAATTCGTTAAG TCACCCAATA 5866
2254 yLysThrThrThrLeuIleLysValPheCysGluThrPheSerLysValA 2271
5867 GCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACATA 5910
2271 snSerLeuIleLeuThrAlaAsnLysSerSerArgGluGluIleLeuAla 2287
5911 AATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCT 5951
2288 LysValAsnArgIleValLeuAspGluGlyAspThrProLeuGlnThrAr 2304
5952 CAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGG 6001
2304 gAspArgIleLeuThrIleAspSerTyrLeuMetAsn AsnArgG 2319
6002 GAATGTATAAAAGGGTTTTG GTGGATGAGGTGCACATGATGCATCAA 6048
2319 lyLeuThrCysLysValLeuTyrLeuAspGluCysPheMetValHisAla 2335
6049 GGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTT 6098
2336 GlyAlaAlaValAlaCysIleGluPheThrLysCysAspSerAlaIleLe 2352
6099 TTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTA 6148
2352 uPheGlyAspSerArgGlnIleHisTyrIleAspArgAsnGlu 2366
6149 GGATGGATTGTGCTGTT TTTGTTCCAAAGAAG 6180
2367 ..LeuAspThrAlaValLeuSerAspLeuAsnArgPheVal AspAsp 2381
6181 GAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTG 6230
2382 GluSerArgValTyrGlyGluValSerTyrArgCysProTrpAspValCy 2398
6231 CTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTG 6280
2398 sAlaTrpLeuSerThrPhe TyrProL 2407
6281 AAAAGGTCGTT AGCGGTAAGGACAAACCAGTA 6312
2407 ysThrValAlaThrThrAsnLeuValSerAlaGlyGlnSer 2420

```
    6313 GTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGA 6362
              |||::::::   |||   |||   ::::::|||||||||
    2421 ......SerMetGlnValArgGluIleGluSerValAspAspValGluTy 2435

    6363 AATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCCGGATA 6412
           :::      |||||||||   |||   |||   ||||||:::|||:
    2435 rSerSerGluPheValTyrLeuThrMetLeuGlnSerGluLysLysAspL 2452

    6413 TG..............AAGAGGTCGTTGAAGGGAAAAGGAAAAGAAACA 6447
         ::              |||||||||   ::::::      |||   |||
    2452 euLeuLysSerPheGlyLysArgSerArgSerSerValGluLysProThr 2468

    6448 CCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGT 6497
           |||:::|||||||||||||||||||||:::|||:::      |||
    2469 ...ValLeuThrValHisGluAlaGlnGlyGluThrTyrArgLysValAs 2484

    6498 ATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGC 6547
          |||   |||||||||       :::|||:::   |||   ::::::   |
    2484 nLeuValArgThrLysPheGlnGluAspAspProPheArgSerGluAsnH 2501

    6548 ATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCT 6597
         |||||   |||:::||||||||||||:::::|||||||:::|||:::::
    2501 isIleThrValAlaLeuSerArgHisValGluSerLeuThrTyrSerVal 2517

    6598 CTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCG 6645
         ||||||||||||   ||||||   ::::::      |||      |||
    2518 LeuSerSerLysArgAspAspAlaIleAlaGlnAlaIleValLysAla 2533

seq_name: sp_virus:Q08534

seq_documentation_block:
ID   Q08534      PRELIMINARY;      PRT;  3074 AA.
AC   Q08534;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (CONTAINS: 66 KDA
DE   PROTEIN).
OS   Sugar beet yellow virus (SBYV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=31714;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UKRAINIAN (BYV-U);
RX   MEDLINE=94082464; PubMed=8259666;
RA   Agranovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
RA   Lunina N.A., Atabekov J.G.;
RT   "Beet yellows closterovirus: complete genome structure and
RT   identification of a leader papain-like thiol protease.";
RL   Virology 198:311-324(1994).
CC   -!- FUNCTION: PROBABLE POLYMERASE.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
CC       THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
DR   EMBL; X73476; CAA51871.1; -.
DR   MEROPS; C42.001; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 2.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   RNA-directed RNA polymerase.
FT   CHAIN         1   3074       348 KDA PROTEIN.
FT   CHAIN         1    588       66 KDA PROTEIN.
FT   MUTAGEN     509    509       C->T: NOT PROCESSED INTO ITS SUBUNITS.
FT   MUTAGEN     517    517       C->T: REDUCED PROCESSING INTO SUBUNITS.
FT   MUTAGEN     518    518       C->T: REDUCED PROCESSING INTO SUBUNITS.
FT   MUTAGEN     556    556       H->E: NO EFFECT.
FT   MUTAGEN     569    569       H->E: NOT PROCESSED INTO ITS SUBUNITS.
FT   MUTAGEN     588    588       G->D: NOT PROCESSED INTO ITS SUBUNITS.
SQ   SEQUENCE   3074 AA;  346028 MW;  D75B9564658BEB0C CRC64;

alignment_scores:
            Quality:  740.50              Length:   2504
              Ratio:  0.647                 Gaps:    120
 Percent Similarity:  45.687   Percent Identity: 22.764

alignment_block:
 US-09-301-906-3 x Q08534   ..

 Align seg 1/1  to: Q08534  from: 1  to: 3074

     409 CCGTATGCCCCCGTCAAGCGCGAGCAGCCTAAACCGGCTGTTAAGCAACA 458
         |||      |||      ||||||         |||:::||||||:::
     319 ProArgLeuProLysPheArgGluValGluIleProValValLysLysHi 335

     459 TGAG..............CAGAAGCCCAAACGGCAAGCGTCACATTGGG 493
                           :::||||||   |||               :
     335 sAlaValProAlaValValSerLysProValArg..............T 347

     494 CTGTTAAGCCAACAGCT...GTTGGCGTCCACGTACCACTTCCTAAAAAA 540
         ::   :::|||:::|||   :::|||::::::            ::::::
     347 hrPheArgProValAlaThrThrGlyAlaGluTyrValAsnAlaArgAsn 363

     541 CAGGAAGCACTGGAGCCAGCGCAATCAGTCCCACAACAGTCGTTGGAGGA 590
         |||   :::   :::|||                ::::::   :::   ::
     364 GlnCysSerArgArgProArg............AsnHisProIleLeuAr 376

     591 GAAGGCCGCCTTGACGTTTGGCCTT..................TTCTTCA 622
         ::::|||:::   |||||||||:::                  ||
     376 gSerAlaSerTyrThrPheGlyPheLysLysMetProLeuGlnArgPheM 393

     623 GTAAAGGTGGGGGTGATGAGAGCGACGCTGTCATCTTGCGGAAAGGGAAA 672
           |||            |||:::            ::::::::::::|||
     393 etLys............GluLysLysGluTyrTyrValLysArgSerLys 405

     673 TTGTTTAAC......AGGGCCCTTAATGTTCCTATTGAT.......... 705
         :::   :::         :::   :::   |||::::::
     406 ValValSerSerCysSerValThrLysSerProLeuGluAlaLeuAlaSe 422

     706 ....GTAAAGAACACGTTCGTTTGGGCT..................AAAA 733
             :::||||||         ::::::                 |||:
     422 rIleLeuLysAsnLeuProGlnTyrSerTyrAsnSerGluArgLeuLysP 439

     734 TCTGGGATGAAGCCTCTCGTAGGAGAGGGTATTTTTACGTCAAAGATAGA 783
         :::::|||                  :::|||         |||
     439 heTyrAsp.....................HisPheIleGlyAspAspPhe 448

     784 GCTGTTAAATTCTTCCCTATTGTGCGGGGTAGGGCTACGATC........ 825
            ::::::      |||:::      |||:::   ::::::
     449 GluIleGluValHisProLeuArgGlyGlyLysLeuSerValLeuLeuIl 465

     826 .............GAGGACTTCATCGTGAATACAGCC......CCAGGGT 856
                    |||   :::   |||   ||||||      |||
     465 eLeuProLysGlyGluAlaTyrCysValValThrAlaAlaThrProGlnT 482

     857 GTGATGTTGCCTTG.....................CCGCGCATT...GAG 882
             :::||||||                     ||||||:::   |||
     482 yrHisAlaAlaLeuThrIleAlaArgGlyAspArgProArgValGlyGlu 498

     883 TTGTGGAGTATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTG 932
         |||   :::   |||                        :::|||   ||
     499 LeuLeuGlnTyrArg.................ProGlyGluGlyLeuCy 509

     933 TTGGTTTAACAATGAGAGGCTGAGGGGAGAAATTTACAGACGTCGTTGCT 982
         |::::::   :::      |||         :::   :::|||
     509 sTyrLeuAlaHisAlaAlaLeuCysCysAlaLeuGlnLysArgThrPheA 526

     983 TCTCATCTTCCTTTTCGATAGGTTTCTTGATGCACCTTGGCTTTAGATCG 1032
           ::::::::::|||   :::|||:::                    :::
     526 rgGluGluAspPhePheValGlyMetTyrProThrLysPheValPheAla 542

    1033 TTAAAGGTCATTAGGTTTGCGGGCACGAACATACTACACATGCCATCACT 1082
```

```
543 LysArgLeuThrGluLysLeuGlyProSerAlaLeuLysHisProValAr 559

1083 CAATGAAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTC..... 1122
559 gGlyArgGlnValSerArgSerLeuPheHisCysAspValAlaSerAlaP 576

1123 .....TATCTCCCCAATGTCCCAAAAACCGCTATCGTC 1155
576 heSerSerProPheTyrSerLeuProArgPhe..... 586

1156 GCTGGCGATAGGACACGGTTGGGAGGG.....GAGAT 1187
587 .....IleGlyGlyValGluGluGluAlaProGluIl 597

1188 CTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGG 1237
597 eThrSerSerLeuLysHis.....LysA 605

1238 TCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTG 1287
605 laIleGluSerValTyrGluArgValSerIleHisLysAspAsnLeuLeu 621

1288 CTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGC 1337
622 AlaArgSerValGluLysAspLeuIleAspPheLysAspGluIleLysSe 638

1338 AATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGC 1387
638 rLeuSerLysGluLysArgSerValThrValProPheTyrMetGlyGluA 655

1388 GGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTC 1437
655 laValGlnSerGlyLeuThrArgAlaTyrProGlnPheAsnLeuSerPhe 671

1438 TCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTG 1487
672 ThrHisSerValTyrSerAspHisProAlaAlaAlaGlySerArgLeuLe 688

1488 TTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATA 1537
688 uGluAsnGluThrLeuAlaSerMetAlaLysSer...SerPheSerAspI 704

1538 TTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTCAT 1587
704 leGlyGlyCysProLeuPheHisIleLysArgGlySerThrAspTyrHis. 720

1588 GTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATCAATGA 1637
721 ValCysArgProIleTyrAspMetLysAspAlaGlnArgArgValSerAr 737

1638 GATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACC 1687
737 g......GluLeuGlnAlaArgGlyLeuValGluAsnLeuSerArgGluG 752

1688 TTCTAACTGAAGCGGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAATCG 1737
752 lnLeuValGluAlaGlnAlaArg...ValSerValCysProHisThrLeu 767

1738 CAGAACTGCGATTCTAGAGCCGATGCGGGTTTTATGGTGGATGTGTACGA 1787
768 GlyAsnCysAsnValLysSerAspValLeuIleMetValGlnValTyrAs 784

1788 TATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTGG 1837
784 pAlaSerLeuAsnGluIleAlaSerAlaMetValLeuLysGluSerLysV 801

1838 TTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTACGGTAACGGT 1887
801 alAlaTyrLeuThrMetValThrProGlyGluLeuLeuAspGluArgGlu 817

1888 GAAGTTTACTTGGAAGAACTC......GATACGTTGGTGAAGAGGGAAGG 1931

818 AlaPheAlaIleAspAlaLeuGlyCysAspValValValAspThrArgAr 834

1932 TGATTACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAACATT 1981
834 gAspMetValGlnTyrLysPheGly......SerSerCysTyrCysHisL 849

1982 CCTTCTCTAACGTAAGCGGGTTT.....TTCACCTTTTCT 2016
849 ysLeuSerAsnIleLysSerIleMetLeuThrProAlaPheThrPhe... 864

2017 TATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGGGATA 2066
865 .....SerGlyAsnLeuPheSerValGluMetTyrGluAs 876

2067 CCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCACCTG 2116
876 nArgMetGlyValAsnTyrTyrLysIleThrArgSerAlaTyrSerProG 893

2117 GA.....ACTGAGGTT 2127
893 luIleArgGlyValLysThrLeuArgTyrArgArgAlaCysThrGluVal 909

2128 ACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCAT 2177
910 ValGlnValLysLeu...ProArgPheAspLysThrLeuLysThrPheLe 925

2178 ACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCGATT 2227
925 uSer.....GlyTyrAspTyrIleTyrLeuAspA 935

2228 CGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACACTATAGGGACA 2277
935 laLysPheValSerArgValPheAspTyrValValSerAsnCysSerVal 951

2278 TTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAA 2327
952 ValAsnSerLysThrPheGluTrpValTrpSerTyrIleLysSerSerLy 968

2328 GACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGATATTT 2377
968 sSerArgValValIleSerGlyLysValIleHisArgAspValHisIleA 985

2378 CTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGGCTCAGGCGATT 2427
985 spLeuLysHisSerGluCysPheAlaAlaValMetLeuAlaValGlyVal 1001

2428 AAGGATAGG.....GCGAAGAGTATTCGCTCCTATAA 2459
1002 ArgSerArgThrThrThrGluPheLeuAlaLysAsnLeuAsnTyrTyrTh 1018

2460 CTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTC. 2508
1018 rGlyAspAlaSerCysPheGluThrIleArgPheLeuPheArgGluTrpS 1035

2509 .....TTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCC 2547
1035 erArgArgAlaTyrAlaGluIleAsnArgSerPheArgLysLeuMetLys 1051

2548 GTCTATGCTAAGGCAATGGTGCACGATAACTTC.....AACGT 2585
1052 SerIleLeuSerAlaGlyLeuAspTyrGluPheLeuAspLeuAspAsnSe 1068

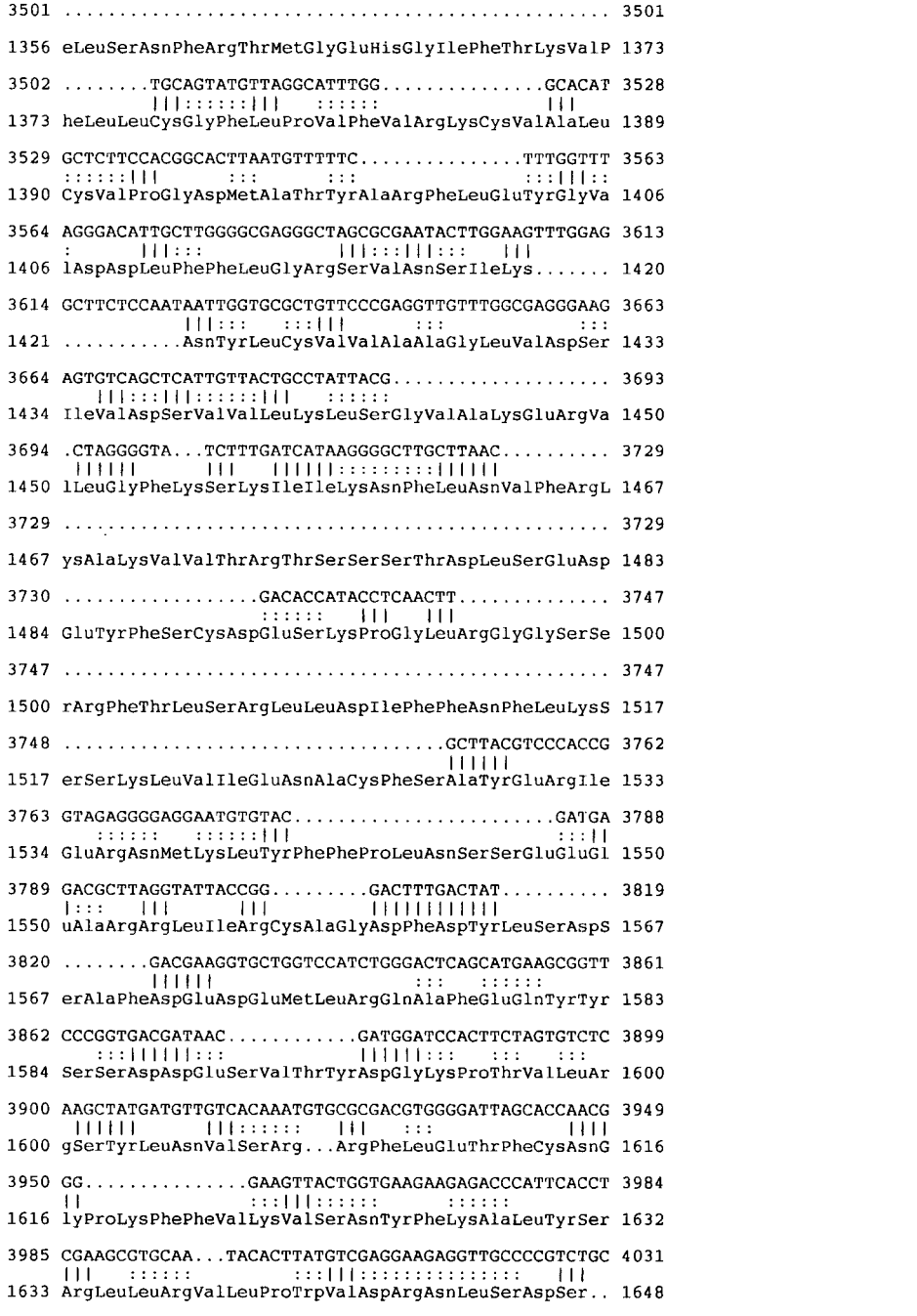
2586 TTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTG 2635
1068 rLeuGlnHisLeuLeuGluTyrSerGluValGluValArg..... 1081

2636 GGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAG 2685
1082 .....ValSerIleAlaGlnAsnGlyGluValAsp..... 1091

2686 CTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAA 2735

1092 .....Cy 1092
```


2736 GAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTA 2785
1092 sAsnGluGluAsnArgValLeuThrGluIleIleAlaGlu..... 1105
2786 AGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACT 2835
1106 AlaAlaAspArgLys..... 1110
2836 GAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTG 2885
1111 SerIleAlaGlnGlyLeuSerGlyAlaLeuSerSerValProTh 1125
2886 T...ACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACC..... 2922
1125 rGlnProArgGlyGlyLeuArgGlyGlySerArgArgSerGlyValSerP 1142
2923 GTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGGG 2961
1142 heLeuTyrAsnLeuValGluGluValGlyAsnLeuPhePheSerValGly 1158
2962 GCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGGTTC 3011
1159 AspAlaValArgPheLeuValLysValPheLysThrPheSerAspSerPr 1175
3012 ACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTGG 3061
1175 oIlePheArgValValArgMetPheLeuAsp.....LeuA 1187
3062 TGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGTCG 3111
1187 laGluAlaAlaSerProPheValSerValValSerLeuCysAlaTrpLeu 1203
3112 CGCGTAGCTGTGAGC.....AAGGTCGTAGC 3137
1204 ArgGluAlaValSerAlaPheSerSerTrpValAlaAspArgThrValSe 1220
3138 TGGAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACCACTT 3187
1220 rGluSerValLysThrPheValAsnArgThrValLysArgPheLeuAsnP 1237
3188 TCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTTGGGT 3237
1237 heMetSerAlaLysThrLeuThrLysLysPhePheArgPhePheLeuSer 1253
3238 GCGGCCCCAATTGCA..... 3252
1254 AlaSerAlaLeuAlaLysThrValValArgLysAlaLysValIleLeuGl 1270
3253 TGGAAG...TATAGGCGCGGAATTGCGGCTAACGCTAGAAGGT 3292
1270 uAlaTyrTrpGluValTrpPheGluSerIleLeuSerAspSerGlyGluT 1287
3293 ACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCC 3342
1287 yr......SerAlaValGluPheCysSerSerValVal..... 1297
3343 GGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGT 3392
1298 IleThrLeuLeuThrAsnSerGlyArgLeuLeu.. 1308
3393 CGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGG..... 3432
1309 ProGlyPheSerProSerAlaIleIleThrGluValLeuLeuA 1323
3433 GCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTT 3477
1323 spLeuAlaThrLysIleSerIleGluValLeuLeuLysGlnIleSerPro 1339
3478 TCTACCTCATACGCCATTTCCGGC..... 3501
1340 AlaAspSerThrAlaSerSerAlaLeuTyrArgArgValLeuSerGluIl 1356


3501 3501
1356 eLeuSerAsnPheArgThrMetGlyGluHisGlyIlePheThrLysValP 1373
3502 TGCAGTATGTTAGGCATTTGG.....GCACAT 3528
1373 heLeuLeuCysGlyPheLeuProValPheValArgLysCysValAlaLeu 1389
3529 GCTCTTCCACGGCACTTAATGTTTTTC.....TTTGGTTT 3563
1390 CysValProGlyAspMetAlaThrTyrAlaArgPheLeuGluTyrGlyVa 1406
3564 AGGGACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTGGAG 3613
1406 lAspAspLeuPhePheLeuGlyArgSerValAsnSerIleLys..... 1420
3614 GCTTCTCCAATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAG 3663
1421 AsnTyrLeuCysValValAlaAlaGlyLeuValAspSer 1433
3664 AGTGTCAGCTCATTGTTACTGCCTATTACG..... 3693
1434 IleValAspSerValValLeuLysLeuSerGlyValAlaLysGluArgVa 1450
3694 .CTAGGGGTA...TCTTTGATCATAAGGGGCTTGCTTAAC..... 3729
1450 lLeuGlyPheLysSerLysIleIleLysAsnPheLeuAsnValPheArgL 1467
3729 3729
1467 ysAlaLysValValThrArgThrSerSerSerThrAspLeuSerGluAsp 1483
3730 GACACCATACCTCAACTT..... 3747
1484 GluTyrPheSerCysAspGluSerLysProGlyLeuArgGlyGlySerSe 1500
3747 3747
1500 rArgPheThrLeuSerArgLeuLeuAspIlePhePheAsnPheLeuLysS 1517
3748 GCTTACGTCCCACCG 3762
1517 erSerLysLeuValIleGluAsnAlaCysPheSerAlaTyrGluArgIle 1533
3763 GTAGAGGGGAGGAATGTGTAC.....GATGA 3788
1534 GluArgAsnMetLysLeuTyrPhePheProLeuAsnSerSerGluGluGl 1550
3789 GACGCTTAGGTATTACCGG.....GACTTTGACTAT..... 3819
1550 uAlaArgArgLeuIleArgCysAlaGlyAspPheAspTyrLeuSerAspS 1567
3820 GACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGCGGTT 3861
1567 erAlaPheAspGluAspGluMetLeuArgGlnAlaPheGluGlnTyrTyr 1583
3862 CCCGGTGACGATAAC.....GATGGATCCACTTCTAGTGTCTC 3899
1584 SerSerAspAspGluSerValThrTyrAspGlyLysProThrValLeuAr 1600
3900 AAGCTATGATGTTGTCACAAATGTGCGCGACGTGGGGATTAGCACCAACG 3949
1600 gSerTyrLeuAsnValSerArg...ArgPheLeuGluThrPheCysAsnG 1616
3950 GG.....GAAGTTACTGGTGAAGAAGAGACCCATTCACCT 3984
1616 lyProLysPhePheValLysValSerAsnTyrPheLysAlaLeuTyrSer 1632
3985 CGAAGCGTGCAA...TACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGC 4031
1633 ArgLeuLeuArgValLeuProTrpValAspArgAsnLeuSerAspSer.. 1648
4032 AGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACG 4081



```
5485 GAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTT 5534
     :::||||||::::::      |||          :::   ||||||
2173 GlnGlyValValSerGlyArgAsn.........SerArgLeuProLeuLy 2186

5535 TCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACG 5584
      |||         :::|||   ::::::   |||:::|||          |
2186 sHisPhe......LysGlyHisGluPheCysPheArgSer......GlyG 2199

5585 GCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATC 5634
     |||||           :::||||||
2199 lyLeu............ValProTyr....................... 2203

5635 TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCT 5684
                                                  |||:::
2204 .....................................AspGlyTh 2206

5685 GTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTTGATAG 5734
           |||   |||      ||||||         |||:::::::
2206 rSerArgValAspThrIle......PheHisThrGlnThrAsnPheValS 2221

5735 GTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCC 5784
     ::|||   :::::         |||   :::      |||:::::::::
2221 erAla...AsnAlaLeuLeuSerGlyTyrLeu......SerTyrArgThr 2234

5785 TTCGAATATAAG.....................TGCTATAATGCTCCCCC 5813
     |||   :::                        |||:::|||||||
2235 PheThrPheThrAsnLeuSerAlaAsnValLeuLeuTyrGluAlaProPr 2251

5814 AGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTAAG......T 5857
     |||||||||||||||||||||||||:::      |||   :::      |
2251 oGlyGlyGlyLysThrThrThrLeuIleLysValPheCysGluThrPheS 2268

5858 CACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAGGAC 5907
     ||      ::::::      :::||||||||||   :::|||   |||:::
2268 erLysValAsnSerLeuIleLeuThrAlaAsnLysSerSerArgGluGlu 2284

5908 ATA..............AATATGGCGGTGAAGAAGAGAGATCCGAATTT 5942
     |||               ::::::::::::   :::   |||      ||
2285 IleLeuAlaLysValAsnArgIleValLeuAspGluGlyAspThrProLe 2301

5943 GGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTA 5992
     |:::      :::         |||:::::::|||   ::::::|||
2301 uGlnThrArgAspArgIleLeuThrIleAspSerTyrLeuMetAsn.... 2316

5993 TCGTCAGGGGAATGTATAAAAGGGTTTTG...GTGGATGAGGTGCACATG 6039
          ||||||:::      :::|||||||   :::||||||      |||
2317 ..AsnArgGlyLeuThrCysLysValLeuTyrLeuAspGluCysPheMet 2332

6040 ATGCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGA 6089
     :::|||   |||               :::      |||   :::::::
2333 ValHisAlaGlyAlaAlaValAlaCysIleGluPheThrLysCysAspSe 2349

6090 AGGCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGA 6139
     ::::::::::||||||||   :::||||||   :::          :
2349 rAlaIleLeuPheGlyAspSerArgGlnIleArgTyr.........GlyA 2363

6140 AGGTGTTTAGGATGGATTGTGCTGTT.................TTTGTT 6171
     ::      ::::::|||   ||||||                 ||||||
2363 rgCysSerGluLeuAspThrAlaValLeuSerAspLeuAsnArgPheVal 2379

6172 CCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAGGTGTCCGTT 6221
           ||||||   ||||||   :::   ||||||||||||||
2380 ...AspAspGluSerArgValTyrGlyGluValSerTyrArgCysProTr 2395

6222 AGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGGAAAAGT 6271
      |||||||||      ||||||::::::
2395 pAspValCysAlaTrpLeuSerThrPhe..................... 2404

6272 GTTACCCTGAAAAGGTCGTT..................AGCGGTAAGGAC 6303
       ||||||:::   |||:::                  :::|||::::::
2405 ..TyrProLysThrValAlaThrThrAsnLeuValSerAlaGlyGlnSer 2420

6304 AAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGA 6353
                     |||::::::   |||   |||   ::::::|||||
2421 ..............SerMetGlnValArgGluIleGluSerValAspAs 2432

6354 CGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGA 6403
     ||||         :::      ||||||||   |||   |||   ||||
2432 pValGluTyrSerSerGluPheValTyrLeuThrMetLeuGlnSerGluL 2449

6404 AGTCGGATATG..............AAGAGGTCGTTGAAGGGAAAAGGA 6438
     ||:::|||:::             |||||||||   ::::::
2449 ysLysAspLeuLeuLysSerPheGlyLysArgSerArgSerSerValGlu 2465

6439 AAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAG 6488
     |||   |||   |||:::|||||||||||||||||||||:::|||:::
2466 LysProThr...ValLeuThrValHisGluAlaGlnGlyGluThrTyrAr 2481

6489 TGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTA 6538
         |||   |||   |||||||||      :::|||:::   |||   :
2481 gLysValAsnLeuValArgThrLysPheGlnGluAspAspProPheArgS 2498

6539 AACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTT 6588
     :::::   ||||||   |||:::|||||||||||||:::::|||||:::
2498 erGluAsnHisIleThrValAlaLeuSerArgHisValGluSerLeuThr 2514

6589 TATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAG 6638
     |||:::::::||||||||||||   ||||||   ::::::      |||
2515 TyrSerValLeuSerSerLysArgAspAspAlaIleAlaGlnAlaIleVa 2531

6639 CGACGCG 6645
         |||
2531 lLysAla 2533
```

```
seq_name: sp_virus:Q66237

seq_documentation_block:
ID   Q66237      PRELIMINARY;      PRT;  3124 AA.
AC   Q66237;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   349-KDA VIRAL POLYPROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T36, QUICK DECLINE;
RX   MEDLINE=95266257; PubMed=7747424;
RA   Karasev A.V., Boyko V.P., Gowda S., Nikolaeva O.V., Hilf M.E.,
RA   Koonin E.V., Niblett C.L., Cline K., Gumpf D.J., Lee R.F.,
RA   Garnsey S.M., Lewandowski D.J., Dawson W.O.;
RT   "Complete sequence of the citrus tristeza virus RNA genome.";
RL   Virology 208:511-520(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T36, QUICK DECLINE;
RA   Karasev A.V.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U16304; AAC59623.1; -.
DR   MEROPS; C42.001; -.
DR   MEROPS; C42.002; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Polyprotein.
SQ   SEQUENCE   3124 AA;  349350 MW;  A02A517D7BDF6C7E CRC64;


alignment_scores:
           Quality:  734.00            Length:   2717
             Ratio:  0.611               Gaps:    111
```

Percent Similarity: 44.203 Percent Identity: 20.537

alignment_block:
US-09-301-906-3 x Q66237 ..

Align seg 1/1 to: Q66237 from: 1 to: 3124

```
    139 TCGGACTTCACACCCAGGTGCGCAGCTATGATTTACATTGGAAAGCTCAC 188
    594 SerAsnPheAspGluLeuCysHisAlaVal......LeuThrArgMetSe 608

    189 CAAAGGGGTGAAGCGTACGTTTGTCCCCCCACCAGTTAAA.......... 228
    608 rLysGlyThrThrLeuGluIleLeuLeuHisSerIleArgGluArgLysG 625

    229 ........GGGTTTGCACGGCAGTACGCTGTTGTCAGCGGCTCAGTCAGC 270
    625 lyLeuLeuGlyPheAlaAlaCysPheArgTrpGlySerSerHisValPhe 641

    271 GCGCTGAGAGGGGAT................................... 285
    642 ValLysArgGlyAsnPheThrAlaArgPheProIleSerAlaValArgTy 658

    286 ..........GGTAAGAAGGTCTTGATGGAGGCAAGGACCTCAACTTCCG 325
    658 rGluAsnGlnTyrAlaAlaValMetMetAspValGlyThrSerTrpValP 675

    326 CAACTTCCGACGTGTCTGATTTCGACGTCGTATTCGAAGCTGTTTCTAAT 375
    675 roLysGlyTyrValProAspValAspValValLysArgArgArgSerArg 691

    376 GCATTACTTGTCGTACACTACCACCGGGTAGTGCCGTATGCCCCCGTCAA 425
    692 AsnArgArgPheProLeuSerHisArgAlaVal................. 702

    426 GCGCGAGCAGCCTAAACCGGCTGTTAAGCAAGATGAGCAGAAGCCCAAAC 475
    703 ....AspGlnSerValProAlaLeu.........GlnSerSerProValG 715

    476 GGCAAGCGTCACATTGGGCTGTTAAGCCAACAGCTGTTGGCGTCCACGTA 525
    715 luValAlaSerValSerSerAspLysProSerSer.............. 726

    526 CCACTTCCTAAAAAACAGGAAGCACTGGAGCCAGCGCAATCAGTCCCACA 575
    727 ............SerGlnLysAlaSerSerProSerThrSerSerSerSe 739

    576 ACAGTCGTTGGAGGAGAAGGCCGCCTTGACGTTTGGCCTTTTCTTCAGTA 625
    739 rSerPheAsnSerThrArgSerSerAspTrpValAsnValGlyPheSerH 756

    626 AA..............................GGTGGGGGTGATGAG 642
    756 isArgThrThrAsnLysValArgGluProArgTyrAlaGlyHisArgArg 772

    643 AGCGACGCTGTCATCTTGCGGAAAGGGAAATTGTTTAAC.......... 681
    773 IleGlyAsnPheThrPheProArgGlyThrValTyrAsnThrProValAs 789

    682 ....AGGGCCCTTAATGTTCCTATTGATGTAAAGAACACGTTCGTTTGGG 727
    789 pGluArgAlaTyrLysArgValLeuArgLeuArgAspThrThrAlaCysS 806

    728 CTAAAATCTGGGATGAAGCCTCTCGTAGGAGAGGGTATTTTTACGTCAAA 777
    806 erPheLeuArgIleLeuLeuAlaArgLeuSerGlyTyrArgProLeuSer 822

    778 GATAGAGCTGTTAAATTCTTC.............................. 798
    823 Asp.........GluPhePheAsnLysCysValThrSerArgPheValAl 836

    799 ..........CCTATTGTGCGGGGT........................ 813
```

```
    836 aCysIleGluProValGluSerGlyLeuValLysValHisPheArgSerA 853

    814 ........AGGGCTACGATCGAGGACTTCATCGTGAATACAGCCCCAGGG 855
    853 spValPheArgAlaSerPheProPheAspGlyGlnProIleHisProAla 869

    856 TGTGATGTTGCCTTGCCGCGCATTGAGTTGTGGAGTATGCGCGAAAGGGC 905
    870 ThrAlaLeuThrLeuGluValValValSerTrpAspTrpMetValThrAl 886

    906 GTTTGTATGCACCACCAAA...GGGTGGTGTTGGTTTAACAAT....... 945
    886 aPheGlnAsnProLeuLysAspGlyTyrCysTyrIleArgHisPheAlaG 903

    945 ................................................... 945

    903 luValSerLeuSerMetGlyArgIlePhePheArgArgAspValAspLeu 919

    946 .....................................GAGAGGCTG.. 954
    920 GlyProPheProTyrValPheGluValGlnHisArgLeuGluArgLeuTy 936

    955 ............................AGGGGAGAAATTTACAGACGTC 976
    936 rGlyLysAlaAlaLeuArgTyrGlyValArgGlyGlnTyrSerAlaProA 953

    977 GTTGCTTCTCATCTTCCTTTTCGATAGGTTTCTTGATGCACCTTGGCTTT 1026
    953 rgCysPheHisCysCysTyrAsn.......................... 960

   1027 AGATCGTTAAAGGTCATTAGGTTTGCGGGCACGAACATACTACACATGCC 1076
    961 ..............................AspSerProArgProMetAl 967

   1077 ATCACTCAATGAAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTCTATC 1126
    967 aSerPheAsnGlyTyrHisLysMetGlyGlyGluAspAsnSerLeuIleI 984

   1127 TCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGTTG 1176
    984 leThrAspThr.....................AspArgLeuArgAla 992

   1177 GGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGT 1226
    993 ValGly.....................................SerAl 996

   1227 CTATTCATCGGTCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGACC 1276
    996 aTyrGluLysValLysArgAsnIleProAspSerLeuLeuValArgSerV 1013

   1277 AATCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCT 1326
   1013 alGluLysAspLeuIleGluPheAsnGlnThrLeuValAspMetHis... 1028

   1327 CAAAGGGACGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTT 1376
   1029 .................ArgSerArgProGlyValValValProPheGl 1039

   1377 TCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTT 1426
   1039 nMetSerGluAsnGlnGlnValMetLeuThrArgAlaTyrProGluPheA 1056

   1427 CGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCC 1476
   1056 snIleAsnPheIleHisSerValHisSerAspHisProValAlaAlaGly 1072

   1477 ATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTG 1526
   1073 SerArgAlaLeuGluAsnHisLeuValArgLysHisAlaGly...ThrAs 1088

   1527 CTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATG 1576
   1088 pTyrSerAspValGlyGlyCysProLeuPheHisLeuArgAlaGlyHisS 1105
```

```
1577 TGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGG 1626
1105 erGlyValHisValCysArgProValTyrAspValLysAspAlaHisArg 1121

1627 AGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTACGT 1676
1122 ArgValValArgHisHisGlnLeuSerLysValSerLeuAsp..... 1135

1677 GTCCAGTGACCTTCTAACTGAAGCG......GCTTCAAAGTCTGTGTCTT 1720
1136 .GlnSerAspGlyValLysGlnValGlyTrpThrValAsnThrAsnSerV 1152

1721 ACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTTT 1770
1152 alCysGlyAsnIleLeuGlyGluCysTyrHisAlaSerGluAlaMetVal 1168

1771 ATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGA 1820
1169 MetValGlnValTyrAspValProLeuArgGluLeuCysArgAlaMetIl 1185

1821 TAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGT 1870
1185 eAsnLysLysThrSerValCysTyrMetThrMetValThrProGlyGluL 1202

1871 TGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTG 1920
1202 euLeuAspAlaArgGluSerPhePheIleLysAspLeuAspCysSerVal 1218

1921 AAGAGGGAAGGTGATTACCTGGCCTACAATGTTGGTCAGTGT......GG 1964
1219 ......GluLeuAspProIleAlaAspArgValValTyrCysPheAsnAs 1233

1965 TGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTTT 2014
1233 nSerAlaTyrThrHisThrTyrSerThrIleCys.............G 1245

2015 CTTATGTACGCACTTCG...........TCCGGGAACGTGTTTAAGCTA 2052
1245 luCysMetArgThrProCysLeuValValAspGlyPheLeuPheThrIle 1261

2053 GAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGC 2102
1262 GluMetValSerLeuArgCysSerValAsnTyrTyrCysIleThrLysSe 1278

2103 TCAGAAGTCACCTGGAACTGAGGTTACG..................... 2130
1278 rSerValCysProArgIleSerGluThrLysArgLeuArgTyrArgArgC 1295

2131 .............TATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATCG 2166
1295 ysAspSerAspLeuIleArgIleLysIleProArgTyrSerSerLysThr 1311

2167 CTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAAT 2216
1312 ArgSerCysLeuPro.....................GlyCysTyrTyrLe 1321

2217 AGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAACA 2266
1321 uTyrLeuAspAlaLysPheValSerArgValTyrGluTyrValValAsnA 1338

2267 CTATAGGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTC 2316
1338 snCysValValValAsnAlaLysThrPheGluTrpThrTrpAsnTyrIle 1354

2317 AGGTCGCAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCAA 2366
1355 LysSerCysLysSerArgValValIleSerGlyLysIleIleHisArgAs 1371

2367 GGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATGG 2416
1371 pGluProIleAlaLeuGluTyrLeuAspGlyPheSerAlaValMetLeuS 1388

2417 CTCAGGCGATTAAGGATAGGGCGAAGAGTATTCGCTCCTATAACTTTATA 2466
1388 erAlaGlyValLysGlyArgGlnAsnAlaAspValSerArgArg...Leu 1403

2467 AAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAGAC 2516
1404 AlaAlaPheSerGlyGluThrSer.............IlePheGluLe 1415

2517 CGTAGGC.....................GATTGTTTT.........T 2533
1415 uValGlyPheAlaIleSerGluLysCysArgAspLeuPheLeuGlyIleG 1432

2534 CGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAAC 2583
1432 lnAspAlaValGluArgCysValLysSerTyrLeuArgAsnSerPheAsn 1448

2584 GTT.......................................... 2586
1449 MetSerPheValAspLeuSerAspProLeuLeuThrIleSerGluTyrSe 1465

2587 ....TTGGAGACGCTTATGTCTATGCCC..................... 2610
1465 rGluLeuAspValProIleAspLeuProGlyPheGlyGlyIleValAspG 1482

2611 ..............................AGAGCGTTCATCCGT 2625
1482 lyAsnGluThrArgLeuMetGluAspGlyValLysSerAlaLeuLeuArg 1498

2626 AAAGTA.....................................CCTGG 2636
1499 LysAlaValGlnAlaGluAlaGlnLeuSerLeuLysArgGlyAsnThrSe 1515

2637 GTCTGTTGTT..................................... 2646
1515 rGluIleValProCysAspAspAspAspAspSerAsnGlyAspAspSerA 1532

2647 ..GTTACCATTTGCACT.............................. 2661
1532 rgValSerLeuAspThrAspTrpTrpPheValLeuArgGlyGlySerSer 1548

2662 ............TCTGGAGCTTCAGACAGGTTG................ 2682
1549 GlyGlyLeuArgAlaGlyAlaAlaProSerLeuPheArgCysLeuLeuAr 1565

2683 .GAGCTCAGGGGTGCCTTTGATATTTCGAAG.................. 2712
1565 gLysLeuGlyGlyGlyLeuCysLeuAlaLysLysValLeuGluLeuLeuV 1582

2713 ..............GAGACCTTCGGTAGG...AAACTGAAGAATAGTCGC 2745
1582 alLysArgValValGluSerPheSerArgLeuLysArgLysThrAlaArg 1598

2746 TTGCGCGTCTTCTCTAGG.............................. 2763
1599 LeuThrSerPheAlaAsnLeuLeuLeuAlaArgLeuSerCysGlyPhePh 1615

2764 ..................GCTATCGTGGAAGATTCAATTAAGGTCATGA 2794
1615 eAspAspThrValPheSerAlaValTrpGluIleAlaIleAsnLeuThrA 1632

2795 AGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCT 2844
1632 rgIleVal.................................AspGly 1636

2845 GTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGC 2894
1637 ValValLysThrLysLeuGlyCysPheArgAsnLeuPheArgLeuIlePr 1653

2895 AGGTCTTCTTGGCGGTTCGAAAGCGACCGTG......GTTTCGAGTGTTT 2938
1653 oLysLeuTyrAlaGlyCysLysAspValValLysAspValSerAsnAlaT 1670

2939 CTAAGGGTTTGGTAGCTCGTGGGGCTGCGACG......AAGGCCTTTTCT 2982
```

```
1670 yrArgSerValValSerPheSerAlaGluValLeuGlyLysSerTyrSer 1686

2983 GGCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTT 3032
1687 LysLeuCysGlnPheCysLeuGlu.....IlePheAspIleGlyLe 1700

3033 AACTGAAGATGAAAGGCTTGAT.....G 3055
1700 uIleGluValSerAlaArgAsnLeuGlyLeuGluThrCysValSerIleA 1717

3056 CTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGC..... 3096
1717 laLeuAlaIlePheAspValLeuIleThrPheProLeuGlySerValSer 1733

3097 .....ATACTGGAGACGTC 3110
1734 PheProValValCysAlaArgValPheGlyArgLeuAlaMetGluValGl 1750

3111 GCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTG 3160
1750 ySerAsnPheLeuPheLysLysValPheGlyLysProGluThrLeuGlyG 1767

3161 AAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTG... 3207
1767 luAspThrPheArgArgThrValSerMetValLeuArgProLysAlaPhe 1783

3207 ..... 3207
1784 AspLysLeuSerValAspAlaLysGlyLeuValArgCysSerGlyValIl 1800

3208 .....CTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTG 3250
1800 eProGlyValLeuArgIleIleLeuSerSerLeuPheSerGluAspAspA 1817

3251 CATGG..... 3255
1817 laTrpValGlyTyrSerLysHisGluValSerSerLeuProIleValThr 1833

3256 .....AAGTATAGGCGCGGAATTGCGGCTAA 3281
1834 PheCysValLysArgLeuSerTyrSerLeuGluArgCysLysGluAlaCy 1850

3282 CGCTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTT 3331
1850 sValAlaPheIleValAspSerAlaLysLysIleLeuLysSerLeuThrS 1867

3332 CACAA.....GCCGCCGGT 3345
1867 erGluPheAsnGluSerIleGlnAsnSerAspLeuThrSerGlyAlaLys 1883

3346 GGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGT 3395
1884 SerLeuTyrGlyHisValSerSerIleLeuTyrThrSerSerGlyValIl 1900

3396 GCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGG..... 3432
1900 eAsnLysAlaGlySerAlaValAlaThrValLysAlaValGlySerAspI 1917

3433 ..GCGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCG..... 3474
1917 leSerGlyArgValLysArgMetArgProArgPheSerLeuSerArgHis 1933

3475 .....TTTTCTACCTCATACGCCATTTCCGGCTGCAGTATGTT 3512
1934 PheAspGluGlnThrSerGluTyrTyrSerAlaSerAspCysSerGluLe 1950

3513 A.....GGCATTTGGGCACATGCTCTTC 3535
1950 uAspGluSerLeuPheSerGluThrProGlyLeuArgGlyAsnAlaArgC 1967

3536 CACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGG 3585
```

```
1967 ysArgAspLeuPhe.....ArgIleValSerArgArg 1977

3586 GCTAGCGCGAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGC 3635
1978 AlaValAla.....CysCy 1982

3636 TGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGC 3685
1982 s.....GluTyrValThrIleLeuLeuLysA 1991

3686 CTATTACGCTAGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACC 3735
1991 rgLeuAlaSerArgAlaGlyGluTyrSerTyrSerTyrValMetAspThr 2007

3736 ATACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGA 3785
2008 LeuValArgLeu...TyrVal.....AlaThrAs 2016

3786 TGAGACGCTTAGGTATTACCGGGACTTTGACTATGAC..... 3822
2016 pAspValValArgArgCysArgSerPheThrPheAspCysThrAspLysA 2033

3823 .....GAAGGT 3828
2033 laSerAsnPheValThrGlyPheLeuSerTrpArgSerLeuArgSerGly 2049

3829 GCTGGTCCATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGA 3878
2050 Ala.....TyrSerAlaThrProGly.....LeuGl 2058

3879 TGGATCCACTTCTAGTGTCTCAAGCTATGATGTTGTCACAAAT...GTGC 3925
2058 yGlyGlySerArgGlyLeuThrSerPheSerHisThrGlyGlyPheThrA 2075

3926 GCGACGTGGGGATTAGCACCAACGGG...GAAGTTACTGGTGAAGAAGAG 3972
2075 rgValIleGlyAlaLeuGlyAsnThrLeuSerValValPheAspGluGlu 2091

3973 .....ACCCATTCACCTCGAAGC..... 3990
2092 PhePheArgThrLeuLeuHisArgLeuArgAlaAspValValValLeuLe 2108

3991 .GTGCAA.....TACACTT 4003
2108 uValGluPheLeuLysAsnLeuProValCysLeuPheSerIleTyrThrP 2125

4004 ATGTCGAG..... 4011
2125 heIleGluCysArgGlyArgIlePheProLeuArgGluPheLeuTrpGly 2141

4012 .....GAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAG 4046
2142 PheCysArgPheValTrpAspValIleGluGlyLeuSerIleAsnThrTy 2158

4047 ACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTG 4096
2158 rGln.....TyrLeuSerAlaValAlaG 2166

4097 AAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGG 4146
2166 luLysValTyrArgGlyAlaPheAsnArgArgPheAsnGluGluAspGlu 2182

4147 GAAACGGCTCGTGAGGTT...GAGGTGGACGGCAAAGGGTTGCTCCCAGA 4193
2183 ArgThrLeuArgValIleThrGluAlaGluGlyArgIleGluPheLeuAr 2199

4194 AAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATG 4243
2199 gValValLeuAlaGlu.....MetGluArgPheArgAlaAlaSerA 2213

4244 GTAACACAGCACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAG 4293
2213 snGluValSer..... 2216
```


4294 TCCAGTCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTC.....AC 4337
2217 SerSerArgSerSerAspSerProSerPheProProTh 2229
4338 CCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGG 4387
2229 rTyrGlyGluIleGluGluLeuAsnAspAspSerSerAspSerAlaLeuV 2246
4388 CG.....ACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAA 4425
2246 alIleGlyValAspAsnGlyThrGlyArgThrIleValAspGlyGluThr 2262
4426 CCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATA...ATTGACAAGGG 4472
2263 AspSerAlaSerGlySerAspSerAspValSerIleProLeuAspArgAs 2279
4473 C..... 4473
2279 pPheGluGlnIleGlyArgLeuGlyGlyAsnGlyAsnValAspValValA 2296
4474 AAGGCCGTCGCTCATGTGGCTGAGAAA 4500
2296 rgLeuGlyLeuLysPheValLeuLysLeuLeuArgAsnLeuPheAspAla 2312
4501 AAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGA 4550
2313 ArgSerPheLysValLeuTyrGlyLeuLeuLeuGlnThrLeuPheTyrGl 2329
4551 AGGCAAGGCCGGTAAACAGCTTTGC.....ATGTTTAGAACGT 4588
2329 yGlySerValAsnLysPheValCysArgLeuValPheSerLeuGlyThrC 2346
4589 GT.....TCCTGCGGTGTG..... 4602
2346 ysLeuProPheGluLeuLeuArgPheAlaCysSerAlaArgArgLeuLeu 2362
4603 CAGCTGGATGTGTACAACGAAGCGAC 4628
2363 LysSerLeuAlaSerPheValTyrGlnThrAsnLeuPheSerGluValMe 2379
4629 TATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAG 4678
2379 tValGluHisLeuPheSerValSerArgLeuLeuTyrAspSer...ArgT 2395
4679 GGAGGAGTGCGGTCTTT..... 4695
2395 yrPheAlaThrValPheSerGluLeuGluArgLeuAsnPheTyrLeuAsp 2411
4696 ...TTCTCAAAGCTGGGTGAGGGGTATACCTAT...AATGGTGGTAGCCA 4739
2412 TyrPheLeuSerLeuProLysArgTyrLeuPheProAsnSerArgAlaAr 2428
4740 TGTTTCATCAGGGTGGCCT..... 4758
2428 gSerCysAlaArgValProValProValIleArgThrGluGlyValAlaP 2445
4759 CGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCA 4797
2445 roValArgGluArgSerLeuAsnAspIleLeuAspSerLeuAspSerLeu 2461
4798 AGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGT 4847
2462 LysValSerAsp.....LeuLysGlyLysSe 2470
4848 ACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCT 4897
2470 rThrPheValAspAspGlyGluSerAspSerGlySerGluHisSerGlyS 2487
4898 TGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAG 4947
2487 erSerIleSerPhe...GlySerValAsnGluGlu..... 2497
4948 GGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTAT 4997
2498 ValLeuAsnAlaValAsnGlyAspSerValCysGl 2509
4998 GCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGAC. 5046
2509 uLysAlaProLeuMetLys...HisValThrAlaLeuSerArgLeuGluA 2525
5047 GAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTC 5088
2525 rgHisProSerGlyArgLysAspIlePro.....ValSerArgLysThr 2539
5089 TTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCG..... 5130
2540 PheValArgGlyGluSerSerGlyValLysSerLysValSerMetCysAs 2556
5131 GATGAGAAGTCACCAGGTGTTCCAAACCAGCAAC 5164
2556 pTyrLeuValThrLeuAsnSerSerPheGlyThrValProAspMetTyrP 2573
5165 CACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAG 5214
2573 ro.....ValAlaArgAspIleThrTyrLysLysLeuThrAsn 2585
5215 GCTCTATCTGAGGGA.....AGTGGTAGGGAAGTCAA 5246
2586 AlaMetArgGluPheTyrTyrSerGlnLysIleThrLeuTyrGluLeuHi 2602
5247 GGGGAGGTCGACATACTCGATATGGTGCGAA.....CAAGATTACG 5287
2602 sGlyLysLeuSer...SerTyrTrpAspGluLeuLysValAlaGlyPheA 2618
5288 TTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAA 5337
2618 spArgLysLeuThrArgMetAspLeuAspAspAspValTyr..... 2631
5338 CCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGA 5387
2631 2631
5388 AGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGA 5437
2632 ValValAspPhe..... 2635
5438 CATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAA 5487
2636 ..AsnArgLysLeuLeuValGlyArgAsn..... 2644
5488 GGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCA 5537
2645 GlySerHi 2647
5538 CGTTTACAAGTCGGGC...ACAGATCTCATTTTTCATTCAACACAAGACG 5584
2647 sProPheArgAsnThrLeuLysSerTyrCysPheMetPheCysAsnAspG 2664
5585 GCTTG.....CGTGTGAGAGACCTACCGTACGTACTCATACCT 5622
2664 lyLeuValProCysProArgSerGlnLysTyrAspTyrAlaLeuIleSer 2680
5623 GAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTT 5672
2681 AsnGlnThrSerPhe.....ValAlaAl 2688
5673 GGGCGACAATCTGTTCGTATGCGAC.....GATATACTGGTTTTCCACG 5716
2688 aAsnSerPheLeuArgAlaValAspGlyLysAspLeuThrPheThrAsnG 2705
5717 ATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTG 5766
2705 luGluHisSerLeuIle..... 2710
5767 GGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGG 5816

```
                                                  |||:::|||||||||||
2711 ..............................ValTyrGluAlaProProGl 2717

5817 TGGCGGTAAGACGACGACGTTAGTGGACGAATTC............GTTA 5854
     |||||||||||||   :::||||||:::::::::             ||||
2717 yGlyGlyLysThrHisSerLeuValAsnSerTyrAlaAspTyrCysValL 2734

5855 AGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGAG 5904
     ||         |||      ::::::|||||||||   :::|||:::
2734 ysVal......SerCysLeuValValThrAlaAsnLysAsnSerGlnThr 2748

5905 GACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAA 5954
     :::|||::::::   :::::::::          :::|||   |||   ::
2749 GluIleSerGlnArgIleSerAsn.........GluLeuMetGlyArgLy 2762

5955 CAGTGCTACC................................ACAGTTA 5971
     :   |||:::                                |||||:
2762 sLeuAlaAlaLysTyrValThrAspAlaAlaSerArgValPheThrValA 2779

5972 ACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTG 6021
     ::|||   :::::::|||   :::   |||         :::   ::::::
2779 spSerTyrLeuMetAsnHisLeu...ArgLeuThrThrGlnLeuLeuPhe 2794

6022 GTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTT 6071
     :::||||||      |||:::|||   |||   :::          |||
2795 IleAspGluCysPheMetValHisAlaGlyAlaIleGlyAlaValValGl 2811

6072 CGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATAC 6121
         |||::::::::::      :::|||||||||||||   :::||||||
2811 uPheThrSerCysLysAlaValValPhePheGlyAspSerLysGlnIleH 2828

6122 CATTCATAAACAGGGAGAAGGTG.........TTTAGGATGGATTGTGCT 6162
       :::|||:::|||:::   :::         |||      |||
2828 isTyrIleHisArgAsnAspLeuGlyValSerPheValAlaAspIleAsp 2844

6163 GTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAG 6212
     :::|||:::   :::   |||      :::|||   :::   ||||||||
2845 AlaPheIle...GlnProGluHisArgIleTyrGlyGluValSerTyrAr 2860

6213 GTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAA 6262
     |||||||   |||:::|||      ||||||::::::
2860 gCysProTrpAspIleCysGluTrpLeuSerGluPhe.............. 2872

6263 CGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTA 6312
                |||||||::   |||:::::::::::::         :::
2873 ...........TyrProArgHisValAlaThrAlaAsnValGlySerIle 2885

6313 GTAAGA...TCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGC 6359
        :::   |||:::|||   :::   |||:::      |||||||||
2886 GlyLysSerSerValSerIleGluGluIleAsnGlyCysAspAspValPr 2902

6360 TGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGG 6409
           :::|||      |||:::      ||||||   |||||||:::|
2902 oTyrAspLysAlaAlaLysTyrIleValTyrThrGlnAlaGluLysAsnA 2919

6410 ATATGAAGAGGTCGTTG............AAGGGAAAAGGAAAAGAAACA 6447
     ||:::::::::   |||               |||::::::|||   :::
2919 spLeuGlnLysHisLeuGlyArgLeuThrValGlyArgAsnLysValVal 2935

6448 CCAGTGATG...ACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGT 6494
     |||::::::   ||||||||L|||:::||||||:::|||:::::::   ||
2936 ProIleValAsnThrValHisGluValGlnGlyGluThrTyrLysArgVa 2952

6495 GGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAAC 6544
     |   |||   |||   |||      :::|||:::   |||::::::::::
2952 lArgLeuValArgPheLysTyrGlnGluAspThrProPheSerSerLysA 2969

6545 CGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCC 6594
       ||||||:::|||:::|||:::|||||||:::   ||||||||||||:::
2969 snHisIleValValAlaLeuThrArgHisValAspSerLeuValTyrSer 2985

6595 GCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGC 6644
     :::|||:::|||:::      |||   :::::|||   |||:::   ||

     2986 ValLeuThrSerArgArgTyrAspAspThrAlaThrAsnIleAspArgAl 3002

     6645 G 6645
          |
     3002 a 3002

seq_name: sp_virus:O40959

seq_documentation_block:
ID   O40959      PRELIMINARY;      PRT;  2301 AA.
AC   O40959;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   METHYLTRANSFERASE (MT) AND HELICASE (HEL) DOMAINS.
OS   Little cherry closterovirus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=53180;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UW2;
RX   MEDLINE=97410293; PubMed=9267008;
RA   Jelkmann W., Fechtner B., Agranovsky A.A.;
RT   "Complete genome structure and phylogenetic analysis of little cherry
RT   virus, a mealybug-transmissible closterovirus.";
RL   J. Gen. Virol. 78:2067-2071(1997).
DR   EMBL; Y10237; CAA71285.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; Methyltransferase; Helicase.
SQ   SEQUENCE   2301 AA;  259625 MW;  2343F291C34E4E45 CRC64;


alignment_scores:
           Quality:  721.00              Length:   2436
             Ratio:  0.663                Gaps:     93
 Percent Similarity:  44.663   Percent Identity: 19.294

alignment_block:
 US-09-301-906-3 x O40959   ..

 Align seg 1/1  to: O40959  from: 1  to: 2301

     202 CGTACGTTTGTCCCCCCACCAGTTAAAGGGTTTGCACGGCAGTACGCTGT 251
         |||||||::      |||         |||:::::::|||
     337 ArgThrLeuArgArgProAlaSerThrGlyTyrValArg........... 349

     252 TGTCAGCGGCTCAGTCAGCGCGCTGAGAGGGGATGGTAAGAAGGTCTTGA 301
                                                 |||   ||||||
     350 ....................................LysProValLeuA 354

     302 TG......GAGGCAAGGACCTCAACTTCCGCAACTTCCGACGTGTCTGAT 345
                 :::|||      :::::::::      |||:::   |||:::
     354 rgAlaHisGlnAlaAlaGlnThrAlaLysIlePheSerSerPheSerGlu 370

     346 TTCGACGTCGTATTCGAA.........GCTGTTTCTAATGCATTACTTGT 386
         |||:::::::::::::|||         |||:::|||:::|||      ::
     371 PheSerIleAlaLeuGluLysGlnLysAlaAlaSerThrAlaSerAlaIl 387

     387 CGTACACTACCACCGGGTAGTGCCGTATGCCCCCGTCAAGCGCGAGCAGC 436
         :                 :::         :::                 ::::::|
     387 e............LysPheGlyAspPheGluValPheAspLeuArgGluP 400

     437 CTAAACCGGCTGTTAAGCAAGATGAGCAGAAGCCCAAACGGCAAGCGTCA 486
         ||      ::::::|||:::      |||                 :::|||
     400 roAspAsnValIleLysSerTyrAlaGln.........AlaLeuThrSer 413

     487 CATTGGGCTGTTAAGCCAACAGCTGTTGGCGTCCACGTACCACTTCCTAA 536
         :::      |||            :::   :::::::::|||:::   ::
     414 AsnLeuProValAspAspArgIleLeuLysLeuAsnIleProValSerSe 430
```

```
 537 AAAACAGGAAGCACTGGAGCCAGCGCAATCAGTCCCACAACAGTCGTTGG 586
 430 rGluValLysThrGluCysProSerThrMetAsnProSerAsnProLysS 447

 587 AGGAGAAGGCCGCCTTGACGTTTGGCCTTTTCTTCAGTAAAGGTGGGGGT 636
 447 erGluIleAlaLysArgAlaValGluLeu...SerSerArgAlaThrSer 462

 637 GATGAGAGCGACGCTGTCATCTTGCGGAAAGGGAAATTGTTTAACAGGGC 686
 463 AspSerSerGlyGluIleTyr.........SerLysThrPheAsnGlnLe 476

 687 CCTTAATGTTCCT..................................... 699
 476 uLeuAsnIleProThrGlyGlyThrIleIleArgHisPheIleAsnGlyV 493

 700 .....................................ATTGATGTA 708
 493 alCysArgSerPheThrLeuAsnArgSerAspGlyArgValPheAsnIle 509

 709 AAGAACACGTTCGTTTGGGCTAAAATCTGGGATGAAGCCTCTCGTAGGAG 758
 510 LeuAsnAsnArgAsnAlaAlaLysIleMetPheAsnLeuSerLeuLysHi 526

 759 AGGGTATTTTTACGTCAAAGATAGAGCTGTTAAATTCTTCCCTATTGTGC 808
 526 sLysLysPheTyrVal................................. 531

 809 GGGGTAGGGCTACGATCGAGGACTTCATCGTGAATACAGCCCCAGGGTGT 858
 532 ...................AspGlyTyrCysArgThrProSerGly... 540

 859 GATGTTGCCTTGCCGCGCATTGAGTTGTGGAGTATGCGCGAAAGGGCGTT 908
 541 .....................................LysLeuPh 543

 909 TGTATGCACCACCAAAGGGTGGTGTTGGTTTAACAATGAGAGGCTGAGGG 958
 543 eSerGluTyrAspAsnGlyPheCysTrpLeuGlnAlaPheAlaMetPheG 560

 959 GAGAAATTTACAGACGTCGTTGC...TTCTCATCTTCCTTTTCGATAGGT 1005
 560 lyLysIleIleProThrPheValGluPheIleProAsnLeuAsnValSer 576

1006 TTCTTGATGCACCTTGGCTTT......AGATCGTTAAAGGTCATTAGGTT 1049
 577 CysLeuLeuGlnAlaGlyLeuProLysSerCysLeuLysTyrMetHisLy 593

1050 TGCGGGCACGAACATACTACACATG...CCATCACTCAATGAAGAGCGTA 1096
 593 sThrGlyAsnAlaLeuTyrHisPheAspProAsn.............. 604

1097 CCTTTGGGTGGAAGGGCGGAGACGTCTATCTCCCCAATGTCCCAAAAACC 1146
 605 ..........LysIleAsnAsnIleTyrHisAlaAsnLeu......... 614

1147 GCTATCGTCGCTGGCGATAGGACACGGTTGGGAGGGGAGATCTTGGCCTC 1196
 615 AsnPheLeuValGlySerAlaMetAsnGluAsnLysAspIleLeuGluSe 631

1197 C.....................................GTCGCCA 1204
 631 rAspAsnIleAspLeuIleValGluGlyPheThrLysGlnValIleAlaL 648

1205 ATGCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGTTTCGAGTATCACC 1254
 648 ysAlaTyrProLysSerAsp......AsnGlnAlaValThrAsnIleLeu 662

1255 AATAGACTGGTATTAAGGGACCAATCGGCATTGCTTTCCCATTTGGACAC 1304
 663 AsnArgValGlyCysArg.......................IleAsnTh 671

1305 GAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAAAAAC 1354
 671 r...TrpCysAsp......ArgLysAsnAspLeuMetIle.......... 681

1355 CCTCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAAGCTG 1404
 682 ..............PheValHisLeuSerThrArgGlnLysLysIleVal 693

1405 AGGGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCAGGAG 1454
 694 CysAspLeuPheProHisLeuLysPheGluPheLysGluThrThrGlnSe 710

1455 TAGTCACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATCTTTT 1504
 710 rSerHisProValPheAsnValValArgThrValSerAsnPheValLeuT 727

1505 CCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTTCACG 1554
 727 yrLysMetGlnGluGlyArgHisPheValAspPheGlyGlyAsnIleGly 743

1555 TATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAGTCCT 1604
 744 ThrValIleAsnSerGluCysAspAspValHisIleCysAsnProValAl 760

1605 AGACGTTAAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTTTCCA 1654
 760 aAspSerArgAspAlaLysArgHisValAspAsnGlyLeuPheLeuAlaL 777

1655 CAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCT 1704
 777 ysSerValGly.................................Val 781

1705 TCAAAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAG 1754
 782 SerAsnAsnIleSerValCysAsnLysLeuAlaGlnHisCysAsnHisLy 798

1755 AGCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGG 1804
 798 sSerAspArgAlaValMetValGluValTyrAspMetThrLeuThrGluM 815

1805 TAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTT 1854
 815 etCysGlnAlaMetLeuAlaHisGlyThrIleArgLeuAspPheIleLeu 831

1855 ATGTTCCCCCTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGA 1904
 832 LeuLeuProGlyAspLeuLeuGluAspPheAsnThrIleThrIlePheAs 848

1905 ACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTG 1954
 848 pGlyGlyCysLysIleThrLysAspAspAspLysValTyrTyrTyrTyrG 865

1955 GTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTT 2004
 865 lyAspAlaAlaGluAlaTyrThrHisAspLeuAsnAsnLeuArgAsnIle 881

2005 TTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGA 2054
 882 MetThrAspAsnLeuVal...CysValAspGlyThrAlaPheLysLysTh 897

2055 GTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTC 2104
 897 rLeuGluThrSerTyrGlyProPheArgHisPheSerLeuThrLysLeuG 914

2105 AGAAGTCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCCGTCGTTC 2154
 914 luThrProSerGlyLysIleGluPheLeuThrMetTyrAspLysCys 930

2155 ...GTGGGCAAATCGCTGGTGTTCATACCTGTT...GTAGCTGGTTCTAG 2198
 931 GluLysAsnLysMetLeuValLysValProMetArgAsnIleAsnGlyMe 947

2199 TGTGTCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCT 2248
```

```
 947 tValThrLeuHisAspLeuIleValAspArgSerMetLeuMetAsnLeuI 964

2249 ATTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAG 2298
 964 leGluTyrCysAlaAsnCysValGluAsnPheAsnArgLysGlyPheGlu 980

2299 TATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAG 2348
 981 HisIleMetSerGlnTyrArgSerArgLysAsnTyrValIleTyrAsnGl 997

2349 TCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTT 2398
 997 yLysValIleHisAspGlnValAspIleProLeuAspLeuLeuProGlyP 1014

2399 TAGTTGTCGCTGTTATGGCTCAGGCGATTAAG......GATAGGGCGAAG 2442
1014 heIleGlyValCysLeuSerGluGlyIleArgLeuAlaGluLysValHis 1030

2443 AGTATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGG 2492
1031 TyrLeuAlaLysTyrSerTyrTyrGln..................... 1039

2493 GGTCTTCAAGCTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAG 2542

1039 ..................................................... 1039

2543 TCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAG 2592
1040 .....TyrTyrAlaProGlyIle.....PheAsnCysMet... 1049

2593 ACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGT 2642
1050 ...IleTyrSerIleArgGlnLeuPheLeuArg.............. 1059

2643 TGTTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGG 2692
1060 .PheLysThrValCysTyrGluLysValLeuAspLeuMetAspPhe.... 1074

2693 GTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGT 2742
1075 ..........LeuPheGlyGluTrpIleThrThrThrAlaArgAsnSer 1087

2743 CGCTTGCGCGTCTTC.........TCTAGGGCTATCGTGGAAGATTCAAT 2783
1088 GlyThrLysIlePheAspIleSerAlaSerValValValLysGlnSerLe 1104

2784 TAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCAATTA 2833
1104 uLeuIle.............SerAlaAspGly.............. 1110

2834 CTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCAC 2883
1111 .....AspValHis 1113

2884 TGTACGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAG 2933

1113 ..................................................... 1113

2934 TGTTTCTAAGGGTTTGGTAGCTCGTGGGGCTGCGACGAAGGCCTTTTCTG 2983

1113 ..................................................... 1113

2984 GCATTACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTA 3033
1114 .....SerIlePhe.............. 1116

3034 ACTGAAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAA 3083
1117 .....ThrGluAsn..... 1119

3084 CTCACCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCG 3133

1119 ..... 1119

3134 TAGCTGGAACGAAAGAATTTTGGAGTGAAGTTTCCTTAAATGACTTCACC 3183
1120 .....PheGlu 1121

3184 ACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTCTTT 3233
1122 LysTyrGlnGlnArg..... 1126

3234 GGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTAACG 3283
1127 .....CysAlaAspL 1130

3284 CTAGAAGGTACGCGGGCAGTAGTTACGAAACTCTAAGCTCGTTAAGTTCA 3333
1130 euGluGlnHisAla...SerGluTyrAspThrAlaSerGluLeuSerAsp 1145

3334 CAAGCCGCCGGTGGTTTACGCGGTTTAACCTCTAGCACAGTATCCGGTGG 3383
1146 AlaIleAspSerGlyGlyMetGlyValSerAspLeuLysIle..... 1159

3384 ATCTTTAGTCGTGCGAAGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGG 3433
1160 .....PheGluLysPheIleThr..... 1165

3434 CGACCGTAGCTAAACGTCAAGTCCCCTTAGCGTTGCTATCGTTTTCTACC 3483
1166 .....LysLeuAlaGluLeuGlyPheGly... 1173

3484 TCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGCACATGCTCT 3533

1173 ..... 1173

3534 TCCACGGCACTTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGA 3583
1174 .GluArgAsnValLeu.....A 1179

3584 GGGCTAGCGCGAATACTTGGAAGTTTGGAGGCTTC.....TCC 3621
1179 rgPheSerAlaLysLeuTyrSerCysPheAspTrpLeuArgArgGlnSer 1195

3622 AATAATTGGTGCGCTGTTCCCGAGGTTGTTTGGCGAGGGAAGAGTGTCAG 3671
1196 AsnAsnAspSerValLeuArgArgValIleGlyTyrPheLysSerLeuPh 1212

3672 CTCATTG......TTACTGCCTATTACGCTAGGGGTATCTTTGATCATAA 3715
1212 eSerLeuSerThrLeuLysAsnIleLeuIleSerProTyrLeuLeuIleL 1229

3716 GG.....GGC 3720
1229 ysTrpIleCysAspCysGlyLysAlaSerPheLysThrPheValSerGly 1245

3721 TTGCTTAAC.....GACACCATACCTCAACTTGCTTA 3752
1246 LeuLeuAsnValPheGlnThrPheSerAspArgIleAsnPheLeuValLy 1262

3753 CGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAGACGCTTAGGTATT 3802
1262 sIleAlaLeuAspSerAspGluAspPheGlnAsnGluValValAlaAlaP 1279

3803 ACCGGGACTTTGAC.....TATGACGAAGGTGCTGGTCCATCTGGG 3843
1279 heArgAsnHisAspLeuGluAlaPheAspAspThrValGlyLysValPhe 1295

3844 ACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCTAG 3893
1296 ThrLysAsnPheAspValValValSerAspAspPheThrLysLysLysIl 1312

3894 TGTCTCA.....AGCTATGAT......GTTGTCACAA 3919
1312 eIleSerIleIleGluAspGluLeuSerTyrGluArgGlyIleAspValA 1329
```


3920 ATGTGCGCGACGTG..... 3933
1329 spValGlnAspValLysLysGluIleLeuGluIleAlaSerSerAlaAsp 1345
3934 GGGATTAGCACCAACGGGGAAGTTACTGGTGAAGA 3968
1346 ThrProThrIleThrLysValAsnAsnLeuGlyHisLeuAspAspAlaLe 1362
3969 AGAGACCCATTCACCTCGAAGCGTGCAATACACTTATGTCGAGGAAGAGG 4018
1362 uGluAsnAspValArgHisLeuValLeuAspIleGlnLeuArgLysGluV 1379
4019 TTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGATCCGTCGGGTTCT 4068
1379 alLeu.....ArgMetValAspGluAspLeuAsnGluAlaLeuGluSer 1393
4069 GGTACCGCTGACGCTATGGCTTTGTTGAAAGTGTG..... 4104
1394 GlyGlyAlaGlyLysSerLysLeuCysAspLysIlePheAsnTyrIleLe 1410
4105 AAAAAAGGTGTCG 4117
1410 uAspArgLeuAsnLeuPheLysHisTyrIleThrSerLysPheGlyTrpA 1427
4118 AC.....GATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTCGTGAG 4161
1427 spPheTyrAspIleLeuHis.....GluSerPheGluGlu 1438
4162 GTTGAGGTGGACGGCAAAGGGTTGCTC.....CCAGAAAGCGTCGTCGG 4205
1439 ValLysValArgLysSerAlaIleLeuCysAlaTyrHisAspAspLeuSe 1455
4206 TGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCAC 4255
1455 rAspIleTyrThr.....AspAsnGlyLeuValL 1465
4256 AAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTG 4305
1465 euLysProLeuAsnLysSerAlaArgGluValLeuTyrAlaArgLeuIle 1481
4306 AGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGC 4355
1482 AlaLys..... 1483
4356 TCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCA 4405
1484 .MetLysPheIleTyrArgAsnPheMetThrThrPheIle..... 1496
4406 CGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACG...ACTCGTGGT 4452
1497 ProSerMetArgPheLeuSerGly 1504
4453 GTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTG.....GCTGA 4496
1505 Ile.....LysAsnLysIleMetSerHisIleCysAspAlaGl 1517
4497 GAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCA 4546
1517 uLysArgGlnAlaIlePheAspLysCysLeuAspAlaThrLeuThrVal. 1533
4547 ATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGC 4596
1534 SerCysThrAla 1537
4597 GGTGTGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTC 4646
1538 ThrAlaGlnLeu.....PheGl 1543
4647 AAACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTT 4696
1543 yGlyAsnPheCysLeuIleArgLeuIle.....LeuSerProIleLeuS 1558
4697 TCTCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCA 4746
1558 erAlaTyrLeuLysLysSerTrpThrTyrTyrPheGlyGluHis..... 1572
4747 TCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCC 4796
1573 AspGlyIleIleSerThrIleIleAlaSe 1582
4797 AAGCGTCTTC.....GACCACTGTTTAGTGCAGAAGTACAAGATGGGTG 4840
1582 rThrLeuLeuGlyGlyGlyAsnTyrSerValAsnLysMetAlaIleValS 1599
4841 GAGGCGTACCATTCCACGCTGATGACGAGGAGTGC..... 4875
1599 erGlySerTyrPheHisPheLysProTyrPheCysAsnLysAlaMetLys 1615
4876 TATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAA 4925
1616 TyrAsnSerPheLysGlnTyrValThr...AspTyrIleAlaArgHisAs 1631
4926 CTTCTCGACTAAGTGCAGGAAG..... 4947
1631 nPhePheThrThrAlaLysThrThrValGlyLeuIleThrSerProValG 1648
4947 4947
1648 lyValAlaValLeuThrValGlyAlaLeuSerGlyIleAsnProPheAla 1664
4948 GGTGG 4952
1665 AlaMetPhePhePheAlaAlaAlaSerHisTyrLysAsnPheTyrAspGl 1681
4953 TAAGGTCATGGTCATAAAC.....GTAGCTTCGGGTGAC.....TATT 4990
1681 yIleValArgValSerAsnIleAlaIleAlaThrGlyAspThrLeuLeuH 1698
4991 TTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCC 5040
1698 isLeuLysProValGlyAlaLeuLysLysAlaIleAsnLysValLysArg 1714
5041 ATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTT 5090
1715 MetLysPheLysAsnSerAsnIleThrSerAsnSerThr...ArgValAs 1730
5091 TGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGT 5140
1730 nGluValAsnAsnAspGlyHisIleAlaAsnAsp...ThrAspAsnLysA 1746
5141 CACCAGGTGTTCCAAAC..... 5157
1746 snSerGluThrSerAsnIleAlaTyrGluGluValGluGluGlnSerAsp 1762
5157 5157
1763 AspAspGluIleValAspIleTyrMetLysArgValLysLysAspLysMe 1779
5158 CAGCAACCACAGAGCCAAGGTGCTACCAGAA 5188
1779 tProIleLysGluTyrIleAspGlnSerGlnGluProGlyValLysGlnA 1796
5189 CAATCACACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGG 5238
1796 laLeuThr.....AsnGlnLeuLeuLysAspSerTyrAsn 1807
5239 GAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGT 5288
1808 AspProAsnAlaSerSerSerTyrSerThrGlyAla..... 1819
5289 TAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAAC 5338
1820 V 1820
5339 CTGACTACACCCCAATG.....ACATTTGAAGTGGTT 5370

```
1820 alAspHisLeuProValAspGlyLysTyrGlnSerValPheAspIleArg 1836

5371 AAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGC 5420
1837 LysAsnHisArgGlyProAspIleSerValGlnLeuCysGluGluLeuPr 1853

5421 TATA...GGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATA 5467
1853 oMetCysGluIleMetArgAsnTyrProAla.....SerLysAsnI 1867

5468 TTGCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTAT 5517
1867 leValPhe......GlnGlnThrValAsnGluValGluAsnSerIleGln 1881

5518 GAATCACTA..................................... 5526
1882 GluPheValGlnIleGluSerLeuThrLeuLeuLeuAsnPheLysLysLe 1898

5527 .....CCGGGCTTTCACGTTTACAAGTCGG 5551
1898 uGluThrThrLeuAspLysPheAspSerGlyMetLysThrTyrArgGlnL 1915

5552 GCACAGATCTCATTTTTCATTCAACACAAGAC.............. 5583
1915 euGlnThrLeuValAsnAspSerThrGlnPheMetTyrValLysGlyGln 1931

5584 ............GGCTTGCGTGTGAGAGACCTACCGTACGTACTCAT 5618
1932 AspTrpValProLeuSerLeuIleHisSerAspValProGlnThrVal.. 1947

5619 AGCTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAG 5668
1948 ....GluTyrSerVal...ThrSerSerArgLysLeuAlaArgPheAspS 1962

5669 CTTTGGGCGACAATCTG...TTCGTATGCGACGATATACTGGTTTTCCAC 5715
1962 erAsnGlyAspGluValGlnPheThrThrThrGluPheAlaAlaGlyPhe 1978

5716 GATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGT 5765
1979 SerAsnLysLysLeuIleSerMetAspLysPheArgGlnSerAlaAsnIl 1995

5766 GGGCGAATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAG 5815
1995 eIleAspSerAspArgLeuSerThrMetLysIleValAsnLysProProG 2012

5816 GTGGCGGTAAGACGACG............ACGTTAGTGGACGAATTCGTT 5853
2012 lySerGlyLysThrThrGlnIleSerGlnMetMetPheAspCysHisIle 2028

5854 AAGTCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGTTCTGA 5903
2029 ArgGlyLysThrCysLeuAlaLeuThrSerThrLysValGlyLysAsnGl 2045

5904 GGACATA............AATATGGCGGTGAAGAAGAGAGATCCGAATT 5941
2045 uLeuIleGluLysLeuLysAsnLeuIleValGlyAlaLysProLysAsnV 2062

5942 TGGAAGGTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTT 5991
2062 alLysThrValAspAla.....PheAsnMetHisGlyLysLysAla 2075

5992 ATCGTCAGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGAT 6041
2076 IleValAspGluLeuTyr...........ValAspGluCysTyrMetSe 2088

6042 GCATCAAGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAG 6091
2088 rHisSerGlyAlaIleLeuMetAlaLeuSerAsnValAspPheThrAsnC 2105

6092 GCCTCTTTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAG 6141
2105 ysLysPhePheGlyAspGluAsnGlnIleProTyrSerValArgLeuThr 2121

6142 GTGTTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGT 6191
2122 GlyPheGluAlaLeuPheAlaThrThrValPheLysTyrCysHisValTh 2138

6192 ATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGT 6241
2138 rHisAspSerIleSerTyrArgCysProAlaAspValCysTyrLeuLeuS 2155

6242 CCTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTT 6291
2155 erHisAlaThrAspArgAlaGlyAsnLysMetTyrThrAsnGlyValTyr 2171

6292 AGCGGTAAGGACAAACCAGTAGTAAGATCGCTGTCCAAAAGGCCAATTGG 6341
2172 AlaValGlyAsnGlnAsn...IleArgThrMetHisLeuAsnAlaLeuTh 2187

6342 AACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGA 6391
2187 rGlyValAlaAsnIleProIleArgGluAspThrValTyrAlaAlaPheT 2204

6392 CCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAAA 6441
2204 hrGlnGlyGluLysAspGluLeuAsnArgPhe...........AlaLys 2216

6442 GAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTGA 6491
2217 LysAsnIleAlaLysThrValAsnGluLeuGlnGlyGlyThrPheLysAs 2233

6492 TGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAAC 6541
2233 nLeuIleValPheArgLeuArgLeuHisAsnAsnValLeuTyrAsnAspT 2250

6542 AACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTAT 6591
2250 hrArgGlnPheIleThrAlaIleSerArgHisThrGluThrPheAsnTyr 2266

6592 GCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGA 6641
2267 PheCysProSerAsnAlaMetAsnAspSerIleSerAsnAlaIleAsnLy 2283

6642 CGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCCGG 6691
2283 sLeuAsn......ThrIleGluAspAsnValValAlaGlnPheGlyPheA 2298

6692 CTGGTTGC 6699
2298 rgGlnCys 2300
```

```
seq_name: sp_virus:Q9WID7

seq_documentation_block:
ID   Q9WID7      PRELIMINARY;      PRT;  3115 AA.
AC   Q9WID7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P349 PROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99190444; PubMed=10092023;
RA   Vives M.C., Rubio L., Lopez C., Navas-Castillo J., Albiach-Marti M.R.,
RA   Dawson W.O., Guerri J., Flores R., Moreno P.;
RT   "The complete genome sequence of the major component of a mild citrus
RT   tristeza virus isolate.";
RL   J. Gen. Virol. 80:811-816(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RA   Moreno P.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y18420; CAA77161.1; -.
DR   INTERPRO; IPR000515; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
SQ   SEQUENCE   3115 AA;  346982 MW;  3EEA612F605D5045 CRC64;


alignment_scores:
            Quality:  712.50              Length:   2350
              Ratio:  0.671                Gaps:     96
 Percent Similarity:  45.191   Percent Identity: 21.234

alignment_block:
 US-09-301-906-3 x Q9WID7   ..

 Align seg 1/1  to: Q9WID7  from: 1  to: 3115

     892 ATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAA 941
         :::|||:::   |||   ||||||   :::   |||      :::|||::
     891 IleArgHisIleAlaGluValCysLeuSerMetGlyLysValPhePheAr 907

     942 CAATGAGAGGCTGAGGGGA.............................. 960
         :::::::         |||
     907 gArgAspValAsnLeuGlyAlaPheProTyrValPheGluValGlnHisA 924

     961 ........GAAATTTACAGACGTCGTTGCTTCTCATCTTCC......... 993
                 ::::::|||   :::|||:::   :::   |||
     924 rgLeuGluArgLeuTyrGlyLysArgAlaValGluTyrSerValArgGly 940

     994 ...TTTTCGATAGGTTTCTTGATGCACCTTGGCTTTAGA...TCGTTAAA 1037
            ||||||         |||   |||      |||      |||   ::
     941 TyrPheSerAlaProArgLeuTyrHisCysCysPheThrAspSerProAr 957

    1038 GGTCATTAGGTTTGCGGGCACGAACATACTACACATGCCATCACTCAATG 1087
         :   :::||||||                                ||||||:
     957 gProPheArgPhe..............................LeuAsnL 964

    1088 AAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTCTATCTCCCCAATGTC 1137
         ::   :::   :::|||   :::   |||   :::   :::   ::::::
     964 ysTyrHisLysMetGlyGlyGluGluGlyAlaLeuLeuIleThrAspThr 980

    1138 CCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGTTGGGAGGGGAGAT 1187
                                  ||||||   |||      |||:::::
     981 .........................AspArgLeuArgAlaValGlySerVa 989

    1188 CTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGG 1237
         :      :::|||   :::||||||   :::                |||   :
     989 lTyrAspLysValLysLysAlaLeuProAsp............SerLeuL 1002

    1238 TCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTG 1287
         ::|||   |||:::   :::   |||:::      :::|||:::
    1002 euValArgSerValGluLysAspLeuIleGluPheAsnGlnThrLeu... 1017

    1288 CTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGC 1337
                 :::||||||:::
    1018 .........ValAspThrGln............................. 1021

    1338 AATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGC 1387
               |||   :::|||:::                  ::::::   :
    1022 .......ArgAlaArgProGlyValAlaValProPheGlnMetSerGluA 1036

    1388 GGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTC 1437
         ::::::::      |||   :::   :::||||||:::::::|||:::|||
    1036 snGlnGlnValMetLeuThrArgAlaTyrProGluPheAsnIleAsnPhe 1052

    1438 TCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTG 1487
         :::   ||||||:::|||:::||||||   |||   :::   |||
    1053 ThrHisSerValHisSerAspHisProValAlaAlaGlySerArgMetLe 1069

    1488 TTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATA 1537
             |||   :::      :::   :::|||:::      :::   |||:
    1069 uGluAsnHisLeuValArgLysTyrAlaGlySer...AspTyrSerAspV 1085

    1538 TTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTCAT 1587
         ::||||||         :::|||::::::|||||||||::::::   |||
    1085 alGlyGlyCysProLeuPheHisLeuArgAlaGlyHisThrGlyValHis 1101

    1588 GTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATC..... 1632
         ||||||:::||||||   ||||||||||||:::   ||||||:::
    1102 ValCysArgProValTyrAspValLysAspAlaHisArgArgValValAr 1118

    1633 .............AATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATT 1669
                      :::::::::   |||      |||   |||:::   :
    1118 gHisHisGlnLeuGlyLysValAlaPheAspHisThrAspGlySerLysG 1135

    1670 CGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCT 1719
         ::                                :::|||   ::::::   :::
    1135 lnVal.......................GlyAlaValAsnThrAsnThr 1143

    1720 TACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTT 1769
            |||::::::         :::|||         ::::::|||
    1144 ValCysGlyAsnValLeuGlyGluCysPheHisAlaSerGluAlaIleVa 1160

    1770 TATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGG 1819
          ||||||:::||||||||:::      ::::::::::::::||||||
    1160 lMetValGlnValTyrAspValProLeuArgGluLeuCysArgAlaMetV 1177

    1820 ATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAG 1869
           :::|||   :::   |||      ::::::::::::   |||   |||
    1177 alAsnLysLysThrSerValCysTyrMetThrMetValThrProGlyGlu 1193

    1870 TTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGT 1919
         ||||||   ::::::   :::   :::::::::||||||||      ||
    1194 LeuLeuAspAlaArgGluSerPhePheIleLysGluLeuAspCysSerVa 1210

    1920 GAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGGTCAGTGT......G 1963
         |      |||   |||   :::|||   :::|||      |||      :
    1210 l......GluLeuAspValValAlaAspArgValValTyrCysPheAsnA 1225

    1964 GTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTT 2013
         :::::   |||   |||::::::|||::::::
    1225 snSerAlaTyrThrHisAlaTyrSerThrIleArg.............. 1236

    2014 TCTTATGTACGCACTTCG............TCCGGGAACGTGTTTAAGCT 2051
         :::|||:::||||||               :::|||   :::|||   ::
    1237 GluTyrMetArgThrProCysValValValAspGlyPheLeuPheThrIl 1253

    2052 AGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGG 2101
         :|||      :::   ||||||:::   ::::::      :::   ::::
    1253 eGluMetValSerLeuArgCysSerValAsnTyrTyrCysIleThrLysS 1270

    2102 CTCAGAAGTCACCTGGA...ACTGAGGTT..................... 2127
         :::::      |||      :::||||||
    1270 erArgValCysProArgIleSerGluValLysArgLeuArgTyrArgArg 1286

    2128 ............ACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATC 2165
                         |||      :::|||   :::   :::|||::
    1287 CysAspSerAspLeuIleArgIleLysIleProArgTyrSerSerLysTh 1303

    2166 GCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAA 2215
         :         :::|||                        :::         :
    1303 rArgSerCysLeuPro.....................GlyCysTyrTyrL 1313

    2216 TAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAAC 2265
         ::   ||||||:::   ||||||:::|||:::|||:::|||::::::|||
    1313 euTyrLeuAspAlaLysPheValSerArgValTyrGluTyrValValAsn 1329

    2266 ACTATAGGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGT 2315
         :::      :::   :::   :::||||||||:::::   :::   ::
    1330 AsnCysValValValValAsnAlaLysThrPheGluTrpThrTrpAsnTyrIl 1346
```



```
1915 gIleArgProSerPheHisLeu..........................  1922

3737 TACCTCAACTTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGAT  3786
                                     :::|||   |||   :::
1923 ..............................AlaArgCysValAsnGlu  1928

3787 GAGACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGGTCC  3836
     :::|||   :::||||||
1929 GlnThrSerGluTyrTyr..............................  1934

3837 ATCTGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGATCCA  3886
                                                    |||:
1935 ............................................SerA  1936

3887 CTTCTAGTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTGGGG  3936
     ::|||:::   |||:::   |||   |||   :::      |||   ||
1936 laSerAspAspSerGluLeuAspGlu.SerLeuLeuSerAspThrProGl  1952

3937 ATTAGCACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCG  3986
     ||||      :::||||||
1952 yLeuArgGlyAsnGlyLys..............................  1958

3987 AAGCGTGCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGCAGCTG  4036
          ||||||   |||:::   |||            |||:::   :::
1959 .....CysAsnAspLeuPhe...ArgValValSerGlyArgValLeuVal  1972

4037 TGGCGGAAAGACAAGGTG......ATCCGTCGGGTTCTGGTACCGCTGAC  4080
     :::|||:::      |||      ::::::|||:::::::::   :::
1973 TyrArgArgArgAlaValGluAlaLeuLysArgLeuIleLeuHisValGl  1989

4081 GCTATGGCTTTTGTTGAAAGTGTGAAAAAAG...GTGTCGACGATGTCTT  4127
          |||      ::::::      :::   |||      :::
1989 yGluCysLeuTyrSerArgIleAlaAspSerPheValArgLeuPheValA  2006

4128 TCACCAACAGTCTAGTGGGGAAA.............................  4150
       ||||||:::   ||||||||
2006 rgThrAsnThrAlaValGlyLysCysGluSerLeuAlaValAspCysAla  2022

4151 ............CGGCTCGTG......AGGTTGAGGTGG...........  4171
                 ||||||         :::   |||
2023 ThrLysIleThrAspLeuValValGlyTyrIleThrTrpArgAsnLeuHi  2039

4172 ...............ACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGG  4205
                    |||:::   |||   |||      :::|||   |
2039 sPheGlyAlaTyrSerThrThrProGlyLeuSer...GlyGlySerArgV  2055

4206 TGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCAC  4255
     ||         ::      |||:::   |||      |||   |||:::
2055 alSerLeuSerArgLeuLeuGlyAspGlyLeuThrCysValIleGlnAsn  2071

4256 AAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTG  4305
               :::      :::      ::::::   :::      |||
2072 PheValThrAlaTrpLeuThrIlePheAsnGlu...ThrPheTyrLeuSe  2087

4306 AGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGC  4355
      :::   ||||||:::         |||:::   |||   :::   :::|
2087 rLeuAlaHisArgMet.........ArgThrAlaSerIleAsnValLeuL  2101

4356 TCAGAAAGAAGTGAAAC.............................  4372
     ||:::         |||
2101 euGluPheLeuLysAsnLeuProValCysLeuValSerIleTyrThrPhe  2117

4373 ............................AAGAAGTACCATTGG.....  4387
                                 |||::::::   |||
2118 IleGluCysArgSerArgIlePheProLeuLysGluPheLeuTrpGlyPh  2134

4387 .............................................  4387

2134 eCysIlePheValTrpAsnIleIleLysGluLeuSerValCysIleTyrG  2151

4388 ......................CGACTGTTT...CGGGCGCCACGCCAA  4411
                           ||||||      ||||||   |||:::
2151 luAsnLeuProValCysIleThrArgLeuCysAsnArgAlaPheArgArg  2167

4412 TCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGAAGATA  4461
     ||||||   ::::::      ::::::|||:::::::::
2168 SerSerAsnGluAspAspGluArgThrLeuGlnValIle...........  2180

4462 ATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAA.........  4501
        |||:::|||:::      :::   :::|||||||:::
2181 ....ThrGluAlaGluAsnArgIleGlyPheLeuArgGluValLeuAlaG  2196

4502 .....AACAGGTACAAGTCGAGC.....................AGCCCA  4525
           ||||||::::::|||                      ||||||
2196 luMetAlaArgTyrArgAlaSerLeuGluIleProValProGlySerPro  2212

4526 AACAG.........................................  4530
     |||
2213 AsnSerProLeuPheProProThrTyrGlyGluIleGluGluLeuGlyAs  2229

4531 ..................AGGAGTTTGACGATCAATGAAGGCAAGGCCG  4561
                        |||   :::   :::::::::|||:::   :
2229 pGlySerSerAspSerAlaArgValIleGlyValAspAsnGlySerAspS  2246

4562 GTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGAT  4611
     ::                     :::   ::      ::::::||||||
2246 er..................MetSerThrGluGlyGluAlaGluLeuAsp  2256

4612 GTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGCATTTACCTT  4661
           :::::::::      :::      :::::::::   |||
2257 SerAspGlyAspSerAspAspSerIleProLeuAspArgAspPheGluGl  2273

4662 TGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTG  4711
      :::   :::|||   |||   :::   |||         :::||||||
2273 nIleGlyGlyLeuGlyGlySerGlyAsnValAspValPheArgLeuGlyL  2290

4712 AGGGGTATACCTATAATGGTGGTAGCCAT..................GTT  4743
           ::::::::::::   :::   :::                   |||
2290 euArgPheValTrpArgMetSerArgAsnLeuPheAsnLeuArgAlaVal  2306

4744 TCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTA  4793
     :::   ::::::      ||||||:::   |||:::::::::   :::
2307 LysPheSerTyrGlyLeuAlaLeuGlnCysIlePheAsnGlyArgSerTh  2323

4794 CCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAG  4843
        :::   :::::::::   |||          :::::::::||||||
2323 rGlyGlnPheIleSerArgLeuLeu.........PheSerPheGlyGlyC  2337

4844 GCGTACCATTCCAC.................................  4857
       :::|||||||:::
2337 ysLeuProPheGluPhePheArgHisValCysAlaValArgTrpLeuSer  2353

4858 ...................................GCTGATGACGAG..  4869
                                        :::|||:::|||
2354 ArgSerLeuAlaLeuHisValArgAsnAlaSerIleCysAspGluGluTh  2370

4869 .............................................  4869

2370 rValGluTrpLeuPheSerPheSerArgPheLeuTyrAspSerArgTyrL  2387

4869 .............................................  4869

2387 euSerLeuValPheSerGluLeuGluArgValGluPheTyrLeuSerTyr  2403

4870 ..................GAGTGCTATCCATCAGATAAC...........  4890
                        :::|||:::      ::::::
2404 LeuPheSerLeuProLysArgCysPheThrHisAsnArgSerValGlnVa  2420

4891 ............CCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACT  4927
                ||||:::   :::|||:::   ||||||   ::::::
2420 lCysValArgAlaProValArgValValArgAsnValGlyLeuValGluA  2437

4928 TCTCGACTAAG.....................................  4938
       :::   |||
2437 spGlyArgLysSerLeuSerGluValLeuAsnSerLeuAspSerLeuLys  2453
```

```
4939 ...TGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGA 4985
2454 ValSerAsnArgLysGlyLysSerAlaValIleGluSerAspSerSerAs 2470

4986 CTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAA 5035
2470 p.....SerGluS 2473

5036 ACTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGC 5085
2473 erGlySerSerSerGlySerIleSer..... 2481

5086 GTCTTTGGTGTA......GGCAGGATGTTGCAGTTAGCC...GGCGGCGT 5126
2482 ...PheGlySerIleAsnGluGluValLeuLysThrAlaAsnGlySerVa 2497

5127 GTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCAAG 5176
2497 lValSerGluLysAla.....SerLeuMetLysHisAlaA 2509

5177 GTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAG..... 5214
2509 laAlaLeuSerGlyLeuGluProAspLeuGlyGlyLysLysAspLeuPro 2525

5215 .....GCTCTATCTGAGGGAAGTGGTAGGGAAGTCAA 5246
2526 IleIleArgArgAsnPheThrArgGlyGluSerSerGlyThrLysSerVa 2542

5247 GGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGT 5296
2542 lGly.....SerLeuC 2546

5297 GTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTAC 5346
2546 ysGluTyrLeuThrThrLeuAsnSerSerPheGlySerValProAspLeu 2562

5347 ACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCCGT 5396
2563 TyrProValAlaArgAspIleThr..... 2570

5397 CGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAGGG 5446
2571 .....TyrLysLys.....MetThrAsnAlaMetArgG 2580

5447 CGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTCTG 5496
2580 luPheTyrTyrSerGlnLysValThrLeuTyrGluLeuHisGlyLysLeu 2596

5497 AAAGTACCTAATCAAGTTTATGAATCACTACCG......GGCTTTCACGT 5540
2597 .....SerSerTyrTrpAspGluLeuLysThrValGlyPheAspAr 2610

5541 TTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGC 5590
2610 gLysLeuAlaLysMetAspLeu.....AspAspAspValG 2622

5591 GTGTGAGAGACCTACCGTAC...GTACTCATAGCTGAAAAAGGTATC... 5634
2622 lyValValAspPheAsnLeuLysValIleValGlyArgHisGlyThrArg 2638

5635 ...TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAA 5681
2639 ProPheArgAsnThrValAsnSerHisGluPheMetPheCysAspAspGl 2655

5682 TCTGTTCGTATGCGAC.....GATATACTGGTTTTCCACG 5716
2655 yLeuValLeuCysProGlyProLysLysArgAspPheAlaLeuIleSerS 2672

5717 ATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTG 5766
2672 erGlnThrGlnPheValAlaAlaAsnSerPheLeuArg...AlaValAsp 2687

5767 GGCGAA......TCGTTTAAGTCCTTCGAATATAAG......TGCTATAA 5804
```

```
2688 GlyLysAspLeuThrPheThrAsnSerGluHisSerLeuIleValTyrGl 2704

5805 TGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATTCGTTA 5854
2704 uAlaProProGlyGlyGlyLysThrHisSerLeuValThrTyrPheAlaA 2721

5855 AG......TCACCCAATAGCACAGCCACCATTACGGCTAATGTGGGAAGT 5898
2721 spTyrCysThrLysValSerCysLeuValValThrAlaAsnLysAsnSer 2737

5899 TCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGG 5948
2738 GlnAlaGluIleSerGlnArgIleSerArgGluPheMetAspArgLysMe 2754

5949 TCTC.....AACAGTGCTACCACAGTTAACT 5974
2754 tLeuAlaLysHisValValLysThrAlaGlyArgValPheThrValAspS 2771

5975 CCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTTGGTG 6024
2771 erTyrLeuMetAsnHisLeu...ArgLeuThrThrAspLeuLeuPheVal 2786

6025 GATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCTTCGC 6074
2787 AspGluCysPheMetValHisAlaGlyAlaIleGlyAlaValValGluPh 2803

6075 AACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATACCAT 6124
2803 eThrSerCysLysAlaValValPhePheGlyAspSerArgGlnIleHisT 2820

6125 TCATAAACAGGGAGAAGGTG.....TTTAGGATGGATTGTGCTGTT 6165
2820 yrIleHisArgAsnAspLeuGlyValSerLeuLeuHisAspIleAspAla 2836

6166 TTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTACAGGTG 6215
2837 PheIle...GluProGlnHisArgIleTyrGlyGluValSerTyrArgCy 2852

6216 TCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGGAACGG 6265
2852 sProTrpAspIleCysGluTrpLeuSer..... 2861

6266 AAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAGTA 6315
2862 ..LysPheTyrProArgHisValAlaThrThrAsnThrGlySerValGly 2877

6316 AGA...TCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTAGCTGA 6362
2878 LysSerSerValSerIleGluValIleAsnGlyCysAspAspValProTy 2894

6363 AATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGATA 6412
2894 rAspSerSerAlaLysTyrIleValTyrThrGlnAlaGluLysAsnGluL 2911

6413 TGAAGAGGTCGTTG.....AAGGGAAAAGGAAAAGAAACACCA 6450
2911 euGlnLysHisLeuGlyArgLeuThrValGlyArgThrLysAlaValPro 2927

6451 GTGATG...ACAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGT 6497
2928 ValValAsnThrValHisGluValGlnGlyGluThrTyrLysLysValAr 2944

6498 ATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGC 6547
2944 gLeuValArgCysLysTyrGlnGluAspThrProPheCysSerAspAsnH 2961

6548 ATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCT 6597
2961 isValValValAlaLeuThrArgHisValAspSerLeuThrTyrSerVal 2977

6598 CTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGCGACGCG 6645
```

```
    2978 LeuAsnSerArgArgTyrAspLysThrAlaSerSerIleAspGluAla 2993

seq_name: sp_virus:Q9IFX0

seq_documentation_block:
ID   Q9IFX0      PRELIMINARY;      PRT;  3115 AA.
AC   Q9IFX0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   P349 PROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T30;
RA   Albiach-Marti M.R., Mawassi M., Gowda S., Satynarayana T., Hilf M.E.,
RA   Shanker S., Almira E.C., Vives M.C., Lopez C., Guerri J., Flores R.,
RA   Moreno P., Garnsey S.M., Dawson W.O.;
RT   "Sequences of Citrus tristeza virus separated in time and space are
RT   essentially identical.";
RL   J. Virol. 74:6856-6865(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T30;
RA   Albiach-Marti M.R., Dawson W.O.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF260651; AAF70348.1; -.
SQ   SEQUENCE   3115 AA;  346874 MW;  6B97F494ED25AD63 CRC64;


alignment_scores:
            Quality:  697.50             Length:   2359
              Ratio:  0.657                Gaps:    100
 Percent Similarity:  44.977   Percent Identity: 21.195

alignment_block:
 US-09-301-906-3 x Q9IFX0   ..

 Align seg 1/1  to: Q9IFX0  from: 1  to: 3115

     892 ATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAA 941
         :::|||:::   |||   ||||||   :::   |||      :::|||::
     891 IleArgHisIleAlaGluValCysLeuSerMetGlyLysValPhePheAr 907

     942 CAATGAGAGGCTGAGGGGA..............................  960
         :::::::         |||
     907 gArgAspValAsnLeuGlyAlaPheProTyrValPheGluValGlnHisA 924

     961 ........GAAATTTACAGACGTCGTTGCTTCTCATCTTCC.........  993
                 :::::::|||   :::|||:::   :::   |||
     924 rgLeuGluArgLeuTyrGlyLysArgAlaValGluTyrSerValArgGly 940

     994 ...TTTTCGATAGGTTTCTTGATGCACCTTGGCTTTAGA...TCGTTAAA 1037
            ||||||         |||   |||      |||      |||   ::
     941 TyrPheSerAlaProArgLeuTyrHisCysCysPheThrAspSerProAr 957

    1038 GGTCATTAGGTTTGCGGGCACGAACATACTACACATGCCATCACTCAATG 1087
         :   :::||||||                                |||||:
     957 gProPheArgPhe.............................LeuAsnL 964

    1088 AAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTCTATCTCCCCAATGTC 1137
         ::   :::   :::|||   :::   |||   :::   :::   ::::::
     964 ysTyrHisLysMetGlyGlyGluGluGlyAlaLeuLeuIleThrAspThr 980

    1138 CCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGTTGGGAGGGGAGAT 1187
                                 ||||||   |||      |||:::::
     981 .......................AspArgLeuArgAlaValGlySerVa 989

    1188 CTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGG 1237
         :      :::|||   :::||||||   :::            |||   :

     989 lTyrAspLysValLysLysAlaLeuProAsp............SerLeuL 1002

    1238 TCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTG 1287
         ::|||   |||:::   :::   |||:::       :::|||:::
    1002 euValArgSerValGluLysAspLeuIleGluPheAsnGlnThrLeu... 1017

    1288 CTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGC 1337
                 :::||||||:::
    1018 .........ValAspThrGln............................. 1021

    1338 AATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGC 1387
                |||   :::|||:::                     ::::::   :
    1022 .......ArgAlaArgProGlyValAlaValProPheGlnMetSerGluA 1036

    1388 GGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTC 1437
         ::::::::      |||   :::   :::||||||::::::|||:::|||
    1036 snGlnGlnValMetLeuThrArgAlaTyrProGluPheAsnIleAsnPhe 1052

    1438 TCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTG 1487
         :::   ||||||:::|||:::||||||   |||   :::   |||
    1053 ThrHisSerValHisSerAspHisProValAlaAlaGlySerArgMetLe 1069

    1488 TTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATA 1537
             |||   :::       :::   :::|||:::       :::   |||:
    1069 uGluAsnHisLeuValArgLysTyrAlaGlySer...AspTyrSerAspV 1085

    1538 TTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATGTGAACTGTCAT 1587
         ::||||||         :::|||::::::|||||||||::::::   |||
    1085 alGlyGlyCysProLeuPheHisLeuArgAlaGlyHisThrGlyIleHis 1101

    1588 GTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCGGAGAATC..... 1632
         ||||||:::||||||   ||||||||||||:::   ||||||:::
    1102 ValCysArgProValTyrAspValLysAspAlaHisArgArgValValAr 1118

    1633 .............AATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATT 1669
                        :::::::::   |||      |||   |||:::   :
    1118 gHisHisGlnLeuGlyLysValAlaPheAspHisThrAspGlySerLysG 1135

    1670 CGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCAAAGTCTGTGTCT 1719
         ::                              :::|||   ::::::   :::
    1135 lnVal.......................GlyAlaValAsnThrAsnThr 1143

    1720 TACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGCCGATGCGGGTTT 1769
            |||::::::         :::|||         :::::|||
    1144 ValCysGlyAsnIleLeuGlyGluCysPheHisAlaSerGluAlaIleVa 1160

    1770 TATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGG 1819
          ||||||:::||||||||:::      :::::::::::::::|||| |
    1160 lMetValGlnValTyrAspValProLeuArgGluLeuCysArgAlaMetV 1177

    1820 ATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAG 1869
           :::|||   :::   |||      ::::::::::::   |||   |||
    1177 alAsnLysLysThrSerValCysTyrMetThrMetValThrProGlyGlu 1193

    1870 TTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGT 1919
         ||||||   ::::::   :::   :::::::::||||||||      ||
    1194 LeuLeuAspAlaArgGluSerPhePheIleLysGluLeuAspCysSerVa 1210

    1920 GAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGGTCAGTGT......G 1963
         |      |||   |||   :::|||   :::|||      |||      :
    1210 l......GluLeuAspValValAlaAspArgValValTyrCysPheAsnA 1225

    1964 GTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTTT 2013
         :::::   |||   |||::::::|||::::::
    1225 snSerAlaTyrThrHisAlaTyrSerThrIleArg............... 1236

    2014 TCTTATGTACGCACTTCG............TCCGGGAACGTGTTTAAGCT 2051
         :::|||:::||||||               :::|||   :::|||   ::
    1237 GluTyrMetArgThrProCysValValValAspGlyPheLeuPheThrIl 1253

    2052 AGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGG 2101
         :|||      :::   ||||||:::   ::::::      :::   ::::
    1253 eGluMetValSerLeuArgCysSerValAsnTyrTyrCysIleThrLysS 1270
```

```
2102 CTCAGAAGTCACCTGGA...ACTGAGGTT..................... 2127
     :::::      |||      :::||||||
1270 erArgValCysProArgIleSerGluValLysArgLeuArgTyrArgArg 1286

2128 ............ACGTATAGGTCGTTGGTCCCGTCGTTCGTGGGCAAATC 2165
                     |||      :::|||   :::   :::|||::
1287 CysAspSerAspLeuIleArgIleLysIleProArgTyrSerSerLysTh 1303

2166 GCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGTCCTTTAAGACAA 2215
     :         :::|||                      :::         :
1303 rArgSerCysLeuPro.....................GlyCysTyrTyrL 1313

2216 TAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCCTACGCGCTCAAC 2265
     ::   ||||||:::   ||||||:::|||:::|||:::|||::::::|||
1313 euTyrLeuAspAlaLysPheValSerArgValTyrGluTyrValValAsn 1329

2266 ACTATAGGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGT 2315
     :::      :::   :::   :::||||||||||:::::   :::   ::
1330 AsnCysValValValAsnAlaLysThrPheGluTrpThrTrpAsnTyrIl 1346

2316 CAGGTCGCAAAAGACCCATGTCATTACAGGGAGTCGCGTTGTCCACAGCA 2365
     ::::|||   |||:::::::|||:::   :::::::::|||:::|||
1346 eLysSerCysLysSerArgValValIleSerGlyLysValIleHisArgA 1363

2366 AGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTATG 2415
       |||   |||:::   :::   :::   |||:::   ::::::::::::
1363 spValProIleAlaLeuGluTyrLeuAspGlyPheSerAlaValMetLeu 1379

2416 GCTCAGGCGATTAAGGATAGGGCGAAGAGTATTCGCTCCTATAACTTTAT 2465
     :::   ::::::|||   |||   ::::::            :::   ::
1380 SerAlaGlyValLysGlyArgGlnAsnAlaGluAlaPheSerArgArgLe 1396

2466 AAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTC...TTTC 2512
     :   |||   :::|||:::   :::::::::|||:::|||.     |||
1396 uAlaAlaPheSerGlyAsp...ThrSerLeuPheGluLeuValArgPheA 1412

2513 AGACCGTAGGCGATTGT.................TTTTCGAACGCAGTC 2544
       :::         |||                  :::::::::||||||
1412 laValSerGluLysCysArgAspLeuPheValGlyIleGlnGluAlaVal 1428

2545 TCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGAC 2594
     :::      :::::::::   :::::::::::|||||||:::      ::
1429 GluArgCysValArgSerTyrLeuArgAsnSerPheAsnMetSerPheVa 1445

2595 GCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTG 2644
     :   :::|||   |||                              :
1445 lAspLeuSerAspPro.................................L 1451

2645 TTGTTACCATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGT 2694
     :::::||||||              |||:::      ||||||
1451 euLeuThrIle..............SerGluTyrSerGluLeuAspVal 1462

2695 GCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCG 2744
        :::|||:::           ||||||          ::::::::::
1463 ProIleAspLeu......ProGlyPheGlyCysIleThrGluGlyAspGl 1477

2745 CTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAG....... 2787
     :   :::                  ::::::||||||:::::|||
1477 uAlaLys..............LeuIleGluAspGlyValLysSerAlaL 1489

2788 ..GTCATGAAGGCAATGAAGACAGAA....................... 2811
       :::   ||||||::::::   |||
1489 euLeuArgLysAlaValArgAspGluAlaGlnLeuIleLeuGluArgGly 1505

2812 ...GATGGAAAACCCCTGCCAATTACTGAAGAT................. 2841
          ::::::   :::|||      :::|||
1506 GlyThrSerGluIleValProCysGlyAspAspAspGlyLeuGlySerAs 1522

2842 ....TCTGTATATGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTA 2887
        |||      ::::::      |||:::   |||:::      ::::
1522 pGlySerAspSerSerLeuGlySerGlySerAspSerSerSerLysAlaS 1539

2888 CGAGGGCAGGTCTTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTT 2937
     ::   :::||||||   :::|||:::      ::::::      |||:::
1539 erSerGlyGlyLeuArgAlaGlyAlaAlaProSerLeuPheArgSerLeu 1555

2938 TCTAAGGGTTTG..............GTAGCTCGT.............. 2958
        |||   |||               |||||||||
1556 LeuLysLysLeuGlyGlyGlyPheCysValAlaArgArgLeuLeuAspLe 1572

2959 ...................GGGGCTGCGACGAAGGCCTTTTCTGGC.... 2985
                         |||:::   ::::::         |||
1572 uLeuCysGluArgLeuValGlySerLeuSerArgLeuLysPheGlyThrT 1589

2986 .....ATTACGTCGTTCTTTTCC..............ACAGGTTCACTA 3015
          :::|||||||||||||||                 :::
1589 hrHisLeuThrSerPhePheSerMetLeuArgAlaLysLeuSerCysAsn 1605

3016 TTCTACGACCGCGGTTTAACTGAAGAT......GAAAGGCTTGATGCTCT 3059
     |||:::|||   ::::::                :::   :::      ||
1606 PhePheAspAspSerIlePheThrAlaValTrpAspThrValMetAsnLe 1622

3060 GGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGT 3109
     |:::|||      |||:::::::::::             ||||||
1622 uIleArgIleValAsnGlyValThrLysThrLys......LeuGluCysP 1637

3110 CGCGCGTAGCT.........GTGAGCAAGGTCGTAGCTGGAACGAAA... 3147
       |||   :::         :::   |||:::   :::|||   |||
1637 heArgAsnValLeuArgLeuIleProLysLeuCysSerGlyCysLysAsp 1653

3147 ..................................................... 3147

1654 PheAlaLysAspPheAlaAspMetCysLysSerValLeuLeuPheSerTh 1670

3148 .GAATTTTGGAGT..................................... 3159
      ||||||:::|||
1670 rGluPhePheSerLysSerTyrThrLysLeuCysArgPheSerLeuGluI 1687

3160 ....................GAAGTTTCCTTAAATGACTTC......... 3180
                         |||:::|||   :::::::::
1687 lePheAspValGlyLeuIleGluIleSerAlaArgAsnMetGlyLeuGlu 1703

3181 ACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCGTC 3230
     |||          :::   :::   ::::::      |||::::::
1704 ThrCysValSerMetAlaSerAlaIlePheAspValIleLeuThrTyrPr 1720

3231 TTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATTGCGGCTA 3280
      ||||||:::      :::                ||||   ::::::|||
1720 oLeuGlySerIleSerPheProValValCysAlaArgValPheGlyAlaL 1737

3281 ACGCTAGAAGGTACGCGGGCAGTAGTTAC..................... 3309
       |||   :::   :::|||||||:::|||
1737 euAlaMetGlu...ValGlySerAsnTyrCysPheLysLysThrLeuGly 1752

3309 ..................................................... 3309

1753 LysProGluIleLeuGlyGluAspAlaPheArgArgThrValSerMetVa 1769

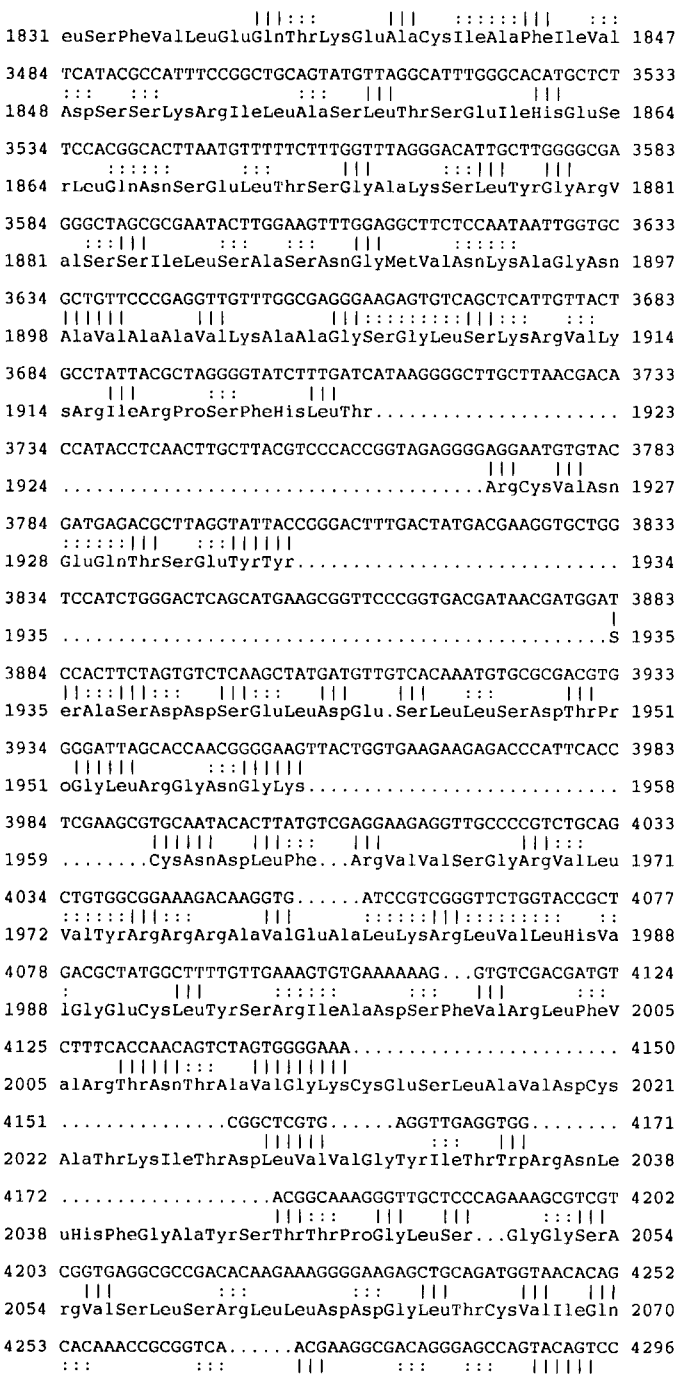
3310 ..........GAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTT 3349
                ::::::::::::::|||||   :::|||   |||   :
1769 lLeuSerAlaLysAlaPheAspLysLeuSerValAspAlaLysGlyLeuV 1786

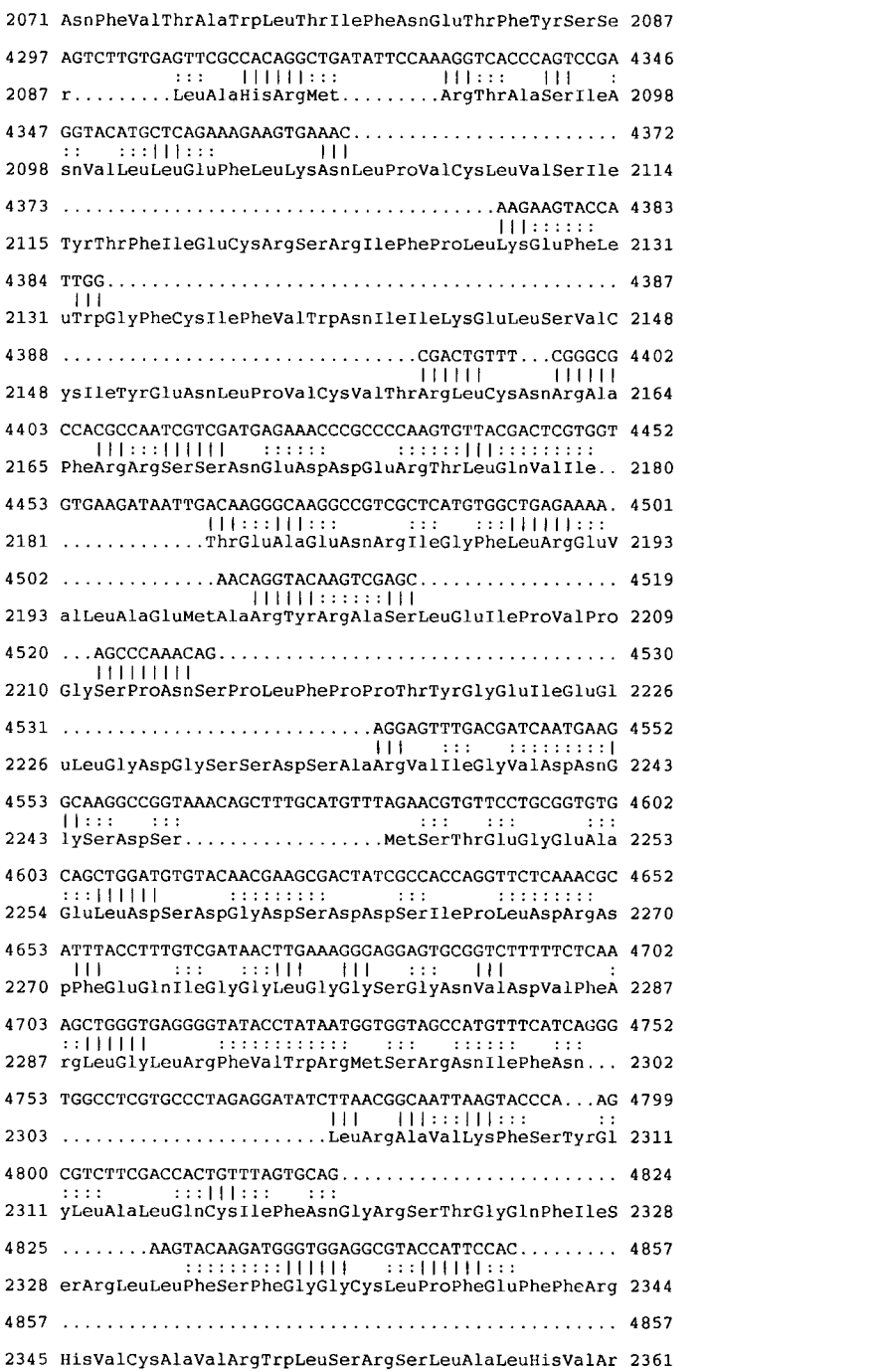
3350 TACGCGGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGA 3399
     ::|||          |||:::   :::   |||          |||:::|||
1786 alArg......CysSerGlyIleIleProGly.........ValLeuArg 1797

3400 AGAGGGTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTA........ 3441
        |||||||||       :::   :::   |||   |||
1798 IleValPheSerSerLeuPheSerAspAlaAspAlaTrpValGlyTyrSe 1814

3442 ....................................GCTAAACGT. 3450
                                        |||||||||
1814 rLysHisGluValSerSerLeuProMetIleThrTyrSerAlaLysArgL 1831

3451 ................CAAGTCCCCTTAGCGTTGCTATCGTTTTCTACC 3483
```


euSerPheValLeuGluGlnThrLysGluAlaCysIleAlaPheIleVal
AspSerSerLysArgIleLeuAlaSerLeuThrSerGluIleHisGluSe
rLeuGlnAsnSerGluLeuThrSerGlyAlaLysSerLeuTyrGlyArgV
alSerSerIleLeuSerAlaSerAsnGlyMetValAsnLysAlaGlyAsn
AlaValAlaAlaValLysAlaAlaGlySerGlyLeuSerLysArgValLy
sArgIleArgProSerPheHisLeuThr
ArgCysValAsn
GluGlnThrSerGluTyrTyr
erAlaSerAspAspSerGluLeuAspGlu.SerLeuLeuSerAspThrPr
oGlyLeuArgGlyAsnGlyLys
.....CysAsnAspLeuPhe...ArgValValSerGlyArgValLeu
lGlyGluCysLeuTyrSerArgIleAlaAspSerPheValArgLeuPheV
alArgThrAsnThrAlaValGlyLysCysGluSerLeuAlaValAspCys
.....CGGCTCGTG.....AGGTTGAGGTGG.....
AlaThrLysIleThrAspLeuValValGlyTyrIleThrTrpArgAsnLe
uHisPheGlyAlaTyrSerThrThrProGlyLeuSer...GlyGlySerA
rgValSerLeuSerArgLeuLeuAspAspGlyLeuThrCysValIleGln
CACAAACCGCGGTCA.....ACGAAGGCGACAGGGAGCCAGTACAGTCC


AsnPheValThrAlaTrpLeuThrIlePheAsnGluThrPheTyrSerSe
r.....LeuAlaHisArgMet.....ArgThrAlaSerIleA
snValLeuLeuGluPheLeuLysAsnLeuProValCysLeuValSerIle
TyrThrPheIleGluCysArgSerArgIlePheProLeuLysGluPheLe
TTGG
uTrpGlyPheCysIlePheValTrpAsnIleIleLysGluLeuSerValC
ysIleTyrGluAsnLeuProValCysValThrArgLeuCysAsnArgAla
PheArgArgSerSerAsnGluAspAspGluArgThrLeuGlnValIle..
GTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAA.
.....ThrGluAlaGluAsnArgIleGlyPheLeuArgGluV
.....AACAGGTACAAGTCGAGC.....
alLeuAlaGluMetAlaArgTyrArgAlaSerLeuGluIleProValPro
...AGCCCAAACAG.....
GlySerProAsnSerProLeuPheProProThrTyrGlyGluIleGluGl
.....AGGAGTTTGACGATCAATGAAG
uLeuGlyAspGlySerSerAspSerAlaArgValIleGlyValAspAsnG
GCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTGTG
lySerAspSer.....MetSerThrGluGlyGluAla
CAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAACGC
GluLeuAspSerAspGlyAspSerAspAspSerIleProLeuAspArgAs
ATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTCAA
pPheGluGlnIleGlyGlyLeuGlyGlySerGlyAsnValAspValPheA
AGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAGGG
rgLeuGlyLeuArgPheValTrpArgMetSerArgAsnIlePheAsn...
TGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCA...AG
.....LeuArgAlaValLysPheSerTyrGl
CGTCTTCGACCACTGTTTAGTGCAG.....
yLeuAlaLeuGlnCysIlePheAsnGlyArgSerThrGlyGlnPheIleS
.....AAGTACAAGATGGGTGGAGGCGTACCATTCCAC.....
erArgLeuLeuPheSerPheGlyGlyCysLeuProPheGluPhePheArg
.....
HisValCysAlaValArgTrpLeuSerArgSerLeuAlaLeuHisValAr

```
4858 .............GCTGATGACGAG........................ 4869
                  :::|||:::|||
2361 gAsnAlaSerIleCysAspGluGluThrValGluTrpLeuPheAsnPheS 2378

4869 ............................................... 4869

2378 erArgPheLeuTyrAspSerArgTyrLeuSerLeuValPheSerGluLeu 2394

4870 ..........................................GAGTG 4874
                                                 :::||
2395 GluArgValGluPheTyrLeuSerTyrLeuPheSerLeuProLysArgCy 2411

4875 CTATCCATCAGATAAC.......................CCTATCTTGA 4900
     |:::      ::::::                      |||:::   :
2411 sPheThrHisAsnArgSerValGlnValCysValArgAlaProValArgV 2428

4901 CGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAG............ 4938
     ::|||:::   ||||||  ::::::   :::   |||
2428 alValArgAspValGlyLeuValGluAspGlyArgLysSerLeuSerGlu 2444

4939 ............................TGCAGGAAGGGTGGTAAGGT 4958
                                 ::::::   ||||||
2445 ValLeuAsnSerLeuAspSerLeuLysValSerAsnArgLysGlyLysSe 2461

4959 CATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTT 5008
       ||||||::      |||:::|||
2461 rAlaValIleGluSerAspSerSerAsp..................... 2470

5009 TTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCATC 5058
                       |||   ::::::   ::::::|||   |||
2471 ..................SerGluSerGlySerSerSerGlySerIle 2480

5059 AGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTA......GGCAG 5102
     |||                          ||||||           ::
2481 Ser.........................PheGlySerIleAsnGluGl 2488

5103 GATGTTGCAGTTAGCC...GGCGGCGTGTCGGATGAGAAGTCACCAGGTG 5149
     ::::|||:::   |||   |||:::|||   :::||||||:::
2488 uValLeuLysThrAlaAsnGlySerValValSerGluLysAla..... 2502

5150 TTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCA 5199
           :::      :::      :::|||         :::   |||
2503 ........SerLeuMetLysHisAlaAlaAlaLeuSerGlyLeuGluPro 2516

5200 AAATCGGGGGGCAAG............................GCTCT 5219
           |||||||||                             :::
2517 AspLeuGlyGlyLysLysAspLeuProIleIleArgArgAsnPheThrAr 2533

5220 ATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATAT 5269
      :::|||:::||||||   :::      |||
2533 gGlyGluSerSerGlyThrLysSerValGly................... 2543

5270 GGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAAT 5319
                              ||||||:::|||   :::   |||
2544 ..................SerLeuCysGluTyrLeuThrThrLeuAsn 2553

5320 CCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGT 5369
           ::::::      ||||||      |||::::::   ::::::::
2554 SerSerPheGlySerValProAspLeuTyrProValAlaArgAspIleTh 2570

5370 TAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGG 5419
     :                                ||| |||
2570 r..............................TyrLysLys....... 2573

5420 CTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATATT 5469
           :::   ::::::   |||   :::      ::::::::::::
2574 ........MetThrAsnAlaMetArgGluPheTyrTyrSerGlnLysVal 2587

5470 GCCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGA 5519
     ::::::         :::|||   |||         ::::::   :::::
2588 ThrLeuTyrGluLeuHisGlyLysLeu.........SerSerTyrTrpAs 2601

5520 ATCACTA......CCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCA 5563
     ::::|||         ||||||            :::      ||||||
2601 pGluLeuLysThrAlaGlyPheAspArgLysLeuAlaLysMetAspLeu. 2617

5564 TTTTTCATTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTAC... 5610
                ::::|||   :::   |||   |||:::
2618 ..............AspAspAspValGlyValValAspPheAsnLeuLys 2629

5611 GTACTCATAGCTGAAAAAGGTATC......TTTACCAAGGGCAAAGATGT 5654
     |||::::::::::::   |||         |||   :::      :::
2630 ValIleValGlyArgHisGlyThrArgProPheArgAsnThrValAsnSe 2646

5655 CGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGCGAC....... 5697
               :::         |||:::|||   :::|||
2646 rHisGluPheMetPheCysAspAspGlyLeuValLeuCysProGlyProL 2663

5698 ........GATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCA 5739
              |||:::   ::::::   :::      :::::::::::|||
2663 ysLysArgAspPheAlaLeuIleSerSerGlnThrGlnPheValAlaAla 2679

5740 CTGAAAGTCGCTCGATGCGGCATGGTGGGCGAA......TCGTTTAAGTC 5783
     :::      |||   ::::::   ||:::      :::|||   ::
2680 AsnSerPheLeuArg...AlaValAspGlyLysAspLeuThrPheThrAs 2695

5784 CTTCGAATATAAG......TGCTATAATGCTCCCCCAGGTGGCGGTAAGA 5827
     :   |||::::::      |||:::||||||||||||||||||||||||
2695 nSerGluHisSerLeuIleValTyrGluAlaProProGlyGlyGlyLysT 2712

5828 CGACGACGTTAGTGGACGAATTCGTTAAG......TCACCCAATAGCACA 5871
     ||   :::||||||      |||:::         :::      |||
2712 hrHisSerLeuValThrTyrPheAlaAspTyrCysThrLysValSerCys 2728

5872 GCCACCATTACGGCTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGT 5921
     :::::|||||||||   :::|||:::   :::|||::::::   ::
2729 LeuValValThrAlaAsnLysAsnSerGlnAlaGluIleSerGlnArgIl 2745

5922 GAAGAAGAGAGATCCGAAT.......................TTGGAAG 5947
     ::::::::::      :::                       ::::::
2745 eSerArgGluLeuMetAspArgLysMetLeuAlaLysHisValIleLysT 2762

5948 GTCTCAACAGTGCTACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTC 5997
          :::   :::   ||||||:::|||   ::::::|||   |||
2762 hrAlaGlyArgValPheThrValAspSerTyrLeuMetAsnHisIle... 2777

5998 AGGGGAATGTATAAAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCA 6047
     |||               ::::::|||||||||      |||:::|||
2778 ArgLeuLysThrAspLeuLeuPheValAspGluCysPheMetValHisAl 2794

6048 AGGCTTACTACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCT 6097
      |||   :::      |||      |||:::::::::      :::|
2794 aGlyAlaIleGlyAlaValValGluPheThrSerCysLysAlaValValP 2811

6098 TTTTTGGAGACATAAATCAGATACCATTCATAAACAGGGAGAAGGTG... 6144
     ||||||||||||   :::|||||||   :::|||:::|||:::   :::
2811 hePheGlyAspSerArgGlnIleHisTyrIleHisArgAsnAspLeuGly 2827

6145 ......TTTAGGATGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGT 6188
           :::      |||      :::|||:::   :::   :::
2828 ValSerLeuLeuHisAspIleAspAlaPheIle...GluProGlnHisAr 2843

6189 TGTATACACTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGT 6238
      :::|||   :::   ||||||||||||||   |||:::|||      |
2843 gIleTyrGlyGluValSerTyrArgCysProTrpAspIleCysGluTrpL 2860

6239 TGTCGTCAATGACCGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTC 6288
     |||||                          |||   ||||||:::   |||
2860 euSer.......................LysPheTyrProArgHisVal 2868

6289 GTTAGCGGTAAGGACAAACCAGTAGTAAGA...TCGCTGTCCAAAAGGCC 6335
     ::::::::::::          |||   :::   |||:::|||   :::
2869 AlaThrAlaAsnThrGlySerValGlyLysSerSerValSerIleGluVa 2885

6336 AATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGT 6385
```

```
       |||:::      |||||||||          ::::::      |||:::
    2885 lIleAsnGlyCysAspAspValProTyrAspSerSerAlaLysTyrIleV 2902

    6386 GCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTG..... 6426
              ||||||   ||||||:::::::::::::::   |||
    2902 alTyrThrGlnAlaGluLysAsnGluLeuGlnLysHisLeuGlyArgLeu 2918

    6427 ...AAGGGAAAAGGAAAAGAAACACCAGTGATG...ACAGTGCATGAAGC 6470
              |||:::   |||   :::||||||:::   ||||||||||||::
    2919 ThrValGlyArgThrLysAlaValProValValAsnThrValHisGluVa 2935

    6471 ACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCG 6520
         :||||||:::|||:::::   |||   |||   |||   |||      :
    2935 lGlnGlyGluThrTyrLysLysValArgLeuValArgCysLysTyrGlnG 2952

    6521 ATGACTCCCTATTCACTAAACAACCGCATATACTTGTTGGTTTGTCGAGA 6570
         ::|||:::   |||   ::::::   |||::::::|||:::|||:::|||
    2952 luAspThrProPheCysSerAspAsnHisValValValAlaLeuThrArg 2968

    6571 CACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAA 6620
         |||:::   ||||||:::|||::::::|||:::|||:::      |||||
    2969 HisValAspSerLeuThrTyrSerValLeuAsnSerArgArgTyrAspLy 2985

    6621 GGTCGGCACATATATTAGCGACGCG 6645
         |:::::::::   |||::::::|||
    2985 sThrAlaSerSerIleAspGluAla 2993

seq_name: sp_virus:O10467

seq_documentation_block:
ID   O10467      PRELIMINARY;      PRT;  3115 AA.
AC   O10467;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   347-KDA POLYPROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SY568;
RA   Yang Z.N., Mathews D.W., Williams A., Dodds J.A., Mirkov T.E.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF001623; AAB57702.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Polyprotein.
SQ   SEQUENCE   3115 AA;  347072 MW;  B406EA0FB38E73F9 CRC64;


alignment_scores:
            Quality:  688.00            Length:   2356
              Ratio:  0.638               Gaps:    100
 Percent Similarity:  45.756   Percent Identity: 20.713

alignment_block:
 US-09-301-906-3 x O10467   ..

 Align seg 1/1  to: O10467  from: 1  to: 3115

     892 ATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAA 941
         :::|||:::   |||   ||||||   :::   |||      :::|||::
     891 IleArgHisIleAlaGluValCysLeuSerMetGlyLysValPhePheAr 907

     942 CAATGAGAGGCTGAGGGGA.............................. 960
         :::::::        |||
     907 gArgAspValAsnLeuGlyAlaProTyrValPheGluValGlnHisA 924

     961 ........GAAATTTACAGACGTCGTTGCTTCTCATCTTCCTTTTCGATA 1002

            ::::::|||   :::|||:::
     924 rgLeuGluArgLeuTyrGlyLysArgAla..................... 933

    1003 GGTTTCTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTTGC 1052
              ::::::   |||   |||:::                |||::
     934 .........ValGluTyrGlyValArgGlyTyr............PheSe 943

    1053 GGGCACGAACATACTACAC...........................ATGC 1075
         ::::|||::::::   |||                          :::
     943 rAlaThrArgLeuTyrHisCysCysPheThrAspSerProArgProPheL 960

    1076 CATCACTCAATGAAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTCTAT 1125
           |||||||||      :::   :::|||   :::   |||   :::
     960 ysSerLeuAsnMetTyrHisLysMetGlyGlyGluGluGlyAlaLeuLeu 976

    1126 CTCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGTT 1175
         :::   ::::::                          ||||||   |||
     977 IleThrAspThr.......................AspArgLeuArgAl 985

    1176 GGGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGG 1225
         |||:::          :::|||   :::   |||   :::
     985 aValGlySerAlaTyrGluLysValLysGlyArgLeuProAsp....... 999

    1226 TCTATTCATCGGTCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGAC 1275
              |||   :::|||   |||:::   :::   |||:::      :::
    1000 .....SerLeuLeuValArgSerValGluLysAspLeuIleGluPheAsn 1014

    1276 CAATCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTC 1325
         |||:::                :::||||||:::
    1015 GlnThrLeu.............ValAspThrGln................. 1021

    1326 TCAAAGGGACGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGT 1375
                           |||   :::|||:::
    1022 ..................ArgAlaArgProGlyValAlaValProPheG 1032

    1376 TTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTT 1425
           ::::::   :::::::::      |||   :::   :::||||||:::
    1032 lnMetSerGluAsnGlnGlnValMetLeuThrArgAlaTyrProGluPhe 1048

    1426 TCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGC 1475
         :::|||:::|||:::   ||||||:::|||:::||||||   |||   ::
    1049 AsnIleAsnPheThrHisSerValHisSerAspHisProValAlaAlaGl 1065

    1476 CATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGT 1525
         :   |||         |||   :::      :::   :::|||:::
    1065 ySerArgMetLeuGluAsnHisLeuValArgLysTyrAlaGlySer...A 1081

    1526 GCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCAT 1575
           :::   |||:::||||||         :::|||::::::||||||||
    1081 spTyrSerAspValGlyGlyCysProLeuPheHisLeuArgAlaGlyHis 1097

    1576 GTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCG 1625
         ::::::   |||||||||:::||||||      :::|||||||:::   ||
    1098 ThrGlyValHisValCysArgProValTyr***IleLysAspAlaHisAr 1114

    1626 GAGAATC.................AATGAGATCCTCTTTCTTTCCACAG 1657
         ||||:::                 :::::::::   |||      |||
    1114 gArgValValArgHisHisGlnLeuSerLysValAlaPheAspHisThrA 1131

    1658 CTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCA 1707
           |||:::   :::                            :::|||
    1131 spGlySerLysGlnVal.......................GlyAlaVal 1139

    1708 AAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGC 1757
         ::::::   :::   |||:::::::         :::|||         ::
    1140 AsnThrAsnThrValCysGlyAsnValLeuGlyGluCysPheHisAlaSe 1156

    1758 CGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAG 1807
         :||||||      ||||||:::||||||||||:::      |||:::::::
    1156 rAspAlaIleValMetValGlnValTyrAspValProLeuGlnGluLeuC 1173

    1808 CAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATG 1857
         :::::||||||   :::|||   |||   |||      ::::::::::::
```

```
1173 ysArgAlaMetValAsnLysLysAlaSerValCysTyrMetThrValVal 1189

1858 TTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACT 1907
     |||   |||||||||   ::::::   :::   :::::::::|||||
1190 ThrProGlyGluLeuLeuAspAlaArgGluSerPhePheIleLysGluLe 1206

1908 CGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGGTC 1957
     ||||      |||      |||   |||   :::|||   :::|||
1206 uAspCysSerVal......GluLeuAspThrValAlaAspArgValValT 1221

1958 AGTGT......GGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGG 2001
       |||      ::::::   |||   |||::::::|||::::::
1221 yrCysPheAsnAsnSerAlaTyrThrHisThrTyrSerThrIleCys... 1236

2002 TTTTTCACCTTTTCTTATGTACGCACTTCG.....TCCGGGAA 2039
                ::::|||:::|||||||          ::::::
1237 .....GluTyrMetArgThrProCysValValValAspSerPh 1249

2040 CGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCA 2089
     :::|||   :::|||      :::   |||||||:::   ::::::
1249 eLeuPheThrIleGluMetValSerLeuArgCysSerValAsnTyrTyrC 1266

2090 CTATGTGTAGGGCTCAGAAGTCACCTGGA.....ACTGAGGTT 2127
       :::   :::::::::     |||                ::::::
1266 ysIleThrLysSerArgValCysProArgIleSerGluIleLysArgLeu 1282

2128 ACGTATAGGTCGTTG.....GTCCCGTCGTT 2153
     ||||||                  :::|||   ::
1283 ArgTyrArgArgCysAspSerAspLeuIleArgIleLysIleProArgTy 1299

2154 CGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGT 2203
     :   :::|||:::          :::|||                  :
1299 rSerSerLysThrArgSerCysLeuPro.....G 1309

2204 CCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCC 2253
     ::         :::   ||||||:::   ||||||:::|||:::|||:::
1309 lyCysTyrTyrLeuTyrLeuAspAlaLysPheValSerArgValTyrGlu 1325

2254 TACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGTATGC 2303
     |||:::::::|||:::      :::   :::   :::||||||||:::::
1326 TyrValValAsnAsnCysValValValAsnAlaLysThrPheGluTrpTh 1342

2304 CGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGTCGCG 2353
     :   :::   ::::::|||   |||::::::|||:::   ::::::::::
1342 rTrpAsnTyrIleLysSerCysLysSerArgValValIleSerGlyLysI 1359

2354 TTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTT 2403
     :::::|||      |||   |||:::   :::   :::   |||:::
1359 leIleHisArgAspValProIleAlaLeuGluTyrLeuAspGlyPheSer 1375

2404 GTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTATTCGCTC 2453
     ::::::::::::::   :::::::|||   |||   ::::::   :::::
1376 AlaValMetLeuSerAlaGlyValLysGlyArg...GlnAsnAlaGluAl 1391

2454 CTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGC 2503
     :::::::          |||:::   ::::::   :::::::::|||:::|
1391 aPheSerArgArgLeuAla***PheSerGlyAsp***SerLeuPheGluL 1408

2504 TCTTC...TTTCAGACCGTAGGCGATTGT.....TTT 2532
     ||      |||                 |||          :::
1408 euValArgPheAlaIleSerGluLysCysArgAspLeuPheLeuGlyIle 1424

2533 TCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAA 2582
     :::::||||||:::      :::|||:::   ::::::::::::|||||
1425 GlnGluAlaValGluArgCysValLysSerTyrLeuArgAsnSerPheAs 1441

2583 CGTTTTGGAGACGCTTATGTCTATGCCCAGAGCGTTCATCCGTAAAGTAC 2632
     |:::      :::   :::|||   |||
1441 nMetSerPheValAspLeuSerAspPro..... 1450

2633 CTGGGTCTGTTGTTGTTACCATTTGCACTTCTGGAGCTTCAGAC...AGG 2679
               ::::::||||||         :::      |||
1451 ..........LeuLeuThrIleSerGluTyrSerGluLeuAspValPro 1463

2680 TTGGAGCTCAGGGGTGCCTTTGATATTTCGAAGGAGACCTTCGGTAGGAA 2729
     :::::::|||   |||         |||::::::|||      ::::::
1464 IleAspLeuProGlyPheGlyCysIleThrGluGluAspGluAlaLys.. 1479

2730 ACTGAAGAATAGTCGCTTGCGCGTCTTCTCTAGGGCTATCGTGGAAGATT 2779
                                              ::::::||||||:
1480 ..................................LeuIleGluAspG 1484

2780 CAATTAAGGTCATGAAGGCAATGAAGACAGAAGATGGAAAACCCCTGCCA 2829
     :::::|||                 |||:::
1484 lyValLysSerAlaLeuLeuArgLysAla..... 1493

2830 ATTACTGAAGATTCTGTATATGCGTTCATAATGGGGAACGTTTCTAAC.. 2877
     :::   :::::::::         :::      |||:::::::|||:::
1494 ValArgAspGluAlaGlnLeuIleLeuGluArgGlyGlyThrSerGluIl 1510

2878 .GTCCACTGTACGAGGGCAGGTCTTCTTGGC..... 2907
      |||   |||               ||||||
1510 eValProCysGlyAspGluGluSerLeuGlyGlyAspGlySerAspSerL 1527

2908 .....GGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTA 2952
          ||||||   :::         |||::::::|||:::|||   :::
1527 euIleGlySerGlySerAspSerProSerAsnAlaSerSerGlyGlyLeu 1543

2953 GCTCGTGGGGCTGCGACGAAGGCCTTT..... 2979
          |||||||||   :::   |||
1544 ArgAlaGlyAlaAlaProSerLeuPheArgSerLeuLeuLysLysLeuGl 1560

2979 ..... 2979

1560 yGlyGlyPheCysMetAlaLysArgLeuLeuAspLeuLeuCysGluArgL 1577

2980 ..................................TCTGGCATTACGTCG 2994
                                        :::   :::||||||
1577 euValAlaSerLeuSerArgLeuLysPheGlyThrThrHisLeuThrSer 1593

2995 TTCTTTTCCACAGGTTCACTA.....TTCTACGAC..... 3024
     :::||||||         :::               |||:::|||
1594 LeuPheSerMetLeuArgIleLysLeuSerArgAsnPhePheAspAspSe 1610

3025 .....C 3025
                                          :
1610 rIlePheThrAlaValTrpAspThrValMetAsnLeuThrArgIleValA 1627

3026 GCGGTTTAACTGAAGATGAA..... 3045
     ::|||:::|||:::   :::
1627 snGlyValThrLysThrLysLeuGluCysPheArgAsnValLeuArgLeu 1643

3045 ..... 3045

1644 IleProLysLeuCysSerGlyCysLysAspLeuAlaLysAspPheAlaAs 1660

3046 .......AGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCA. 3087
            ||||||      |||:::   ||||||         :::|||
1660 pMetCysArgLeuValPheLeuLeuSerThrGluPhePheSerLysSerT 1677

3088 ..................................CCGGTGGGCATACTG 3102
                                              :::|||::::::
1677 yrAlaLysLeuCysArgPheSerLeuGluIlePheAspIleGlyLeuIle 1693

3103 GAGACGTCG.....CGCGTAGCTGTGAGC..... 3126
     |||   |||               :::::::::||||||
1694 GluIleSerAlaArgAsnMetGlyLeuGluThrCysValSerMetAlaSe 1710

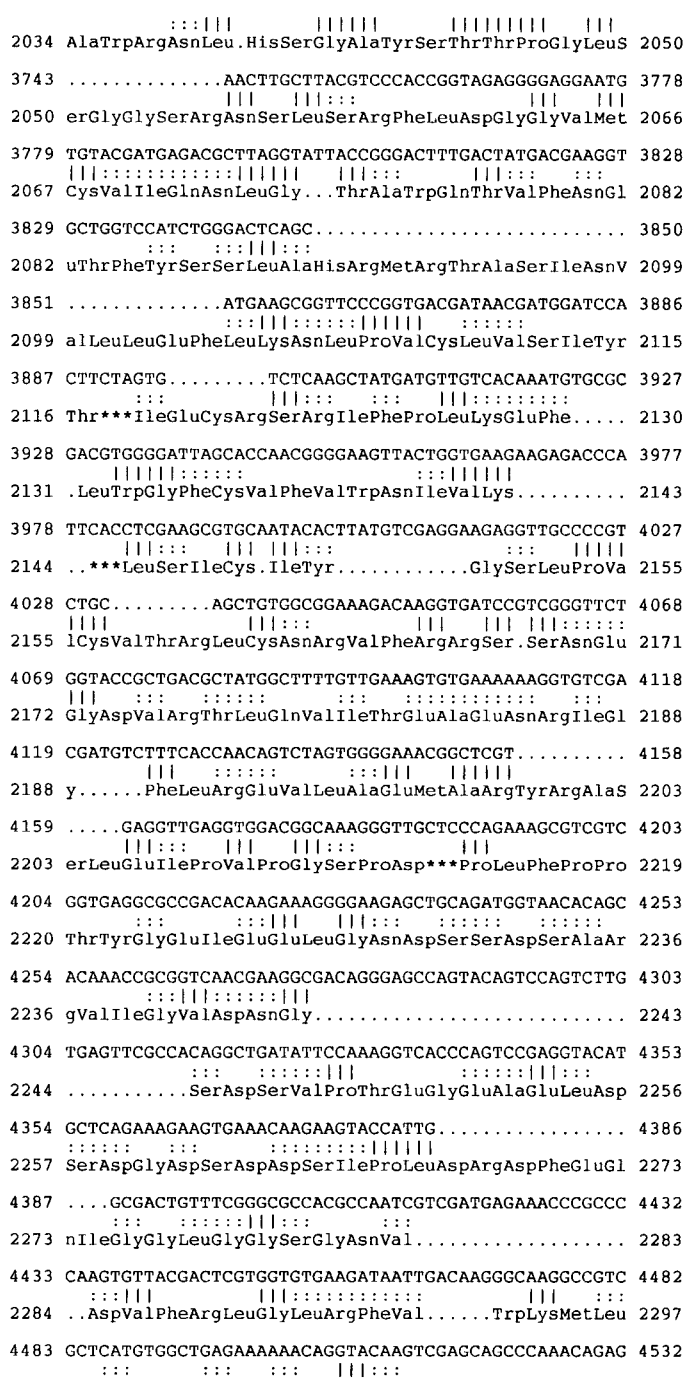
3126 ..... 3126

1710 rAlaIlePheAspValIleLeuThrTyrProLeuGlySerIleSerPheP 1727

3126 ..... 3126

1727 roIleValCysAlaArgValPheGlyAlaLeuAlaMetGluValGlySer 1743
```





```
2298 ArgAsnSerPheAsnLeuArgAlaValLysPhePheTyrGlyLeuAlaLe 2314

4533 GAGTTTGACGATCAATGAAGGCAAGGCCGGT..................... 4563
      :::   :::::::|||   |||:::   |||
2314 uGlnCys***PheAsnGlyGlySerMetGlyGlnPheThrSerArgLeuL 2331

4564 .....................................AAACAGCTTTGC 4575
                                          :::::::::|||
2331 euPheSerPheGlyAsnCysLeuProPheGluPhePheArgHisValCys 2347

4576 ATGTTTAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAAC..... 4620
          |||       |||   ::::::   |||  |||   |||
2348 SerValArgTrpLeuSerLysSerLeuAlaLeuArgValArgAsnAlaSe 2364

4621 ..........GAAGCGACTATCGCC..............ACCAGGTTC. 4644
                |||   |||:::                 :::||||||
2364 rIleCysAspGluGluThrValGluArgPhePheSerPheSerArgPheL 2381

4645 ....................TCAAACGCATTTACCTTTGTCGATAACTTG 4674
                          |||:::::::|||:::   ::::::   :::
2381 euTyrAspSerArgTyrLeuSerAspValPheSerValLeuGlu***Val 2397

4675 AAAGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTA 4724
               :::   :::||||||        :::      |||
2398 AspPheTyrLeuSerTyrLeuPheSerLeuProLysArgCysLeuThrAr 2414

4725 TAATGGTGGTAGCCATGTTTCATCAGGGTGGCCT...CGTGCCCTAGAGG 4771
      |||   :::   :::|||   :::      |||   |||:::::::::|
2414 gAsnArgSerValArgValCysAlaArgAlaProValArgValIleArgA 2431

4772 ATATCTTAACGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTG 4821
     ||:::                                      ||||||
2431 spVal.....................................GlyLeuVal 2435

4822 CAGAAGTACAAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGA 4871
     :::      |||   :::
2436 GluAspGlyLysLysSer.............................. 2441

4872 GTGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGG 4921
                              ||||||      |||   :::
2442 ........................LeuThrGluValLeuAsnSerLeuA 2450

4922 CAAACTTCTCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAAC 4971
       ::::::::::::::::   |||||||||   ::::::::::::
2450 spSerLeuLysValSerAspArgLysGlyLysSerAlaLeuIle...... 2464

4972 GTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCA 5021

2464 ................................................. 2464

5022 CTTGCATTCAGTAAACTCCATCGACGAA...GGGCGCATC......AGTT 5062
        :::|||   ||||||   :::|||   ||||||         |||:
2465 ....GluSerAspAsnSerAspSerGluSerGlyArgSerPheGlySerI 2480

5063 TGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAG 5112
     :::::|||   ::::::::::::                :::::::|||:::
2480 leSerPheGlySerValAsnGlu..............GluValLeuGlu 2491

5113 ...TTAGCCGGCGGCGTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCA 5159
        :::   |||||||||   :::||||||:::|||
2492 AlaValAsnGlyGlyValValSerGluLysAlaPro............Le 2504

5160 GCAACCACAGAGCCAAGGTGCTACCAGAACAATCACACCAAAATCGGGGG 5209
      :::   ::::::   :::   :::|||                ||||||:
2504 uMetLysHisAlaAlaAlaLeuSerArgLeuGluProAspLeuSerGlyA 2521

5210 GCAAGGCTCTA.........................TCTGAGGGAAGT 5232
     ::|||   |||                        :::|||:::|||
2521 snLysAspLeuProIleValArgArgSerPheThrArgGlyGluSerSer 2537

5233 GGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGA 5282
     |||   :::      |||
2538 GlyThrLysProAlaGly............................... 2543

5283 TTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTC 5332
                  |||   :::|||   :::   |||      ::::::
2544 .......SerLeuCysValTyrLeuThrThrLeuAsnSerSerPheGlyS 2558

5333 TTGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACC 5382
          ||||||      |||:::::   :::::::::
2558 erValProAspLeuTyrProValAlaArgAspIleThr............ 2570

5383 TCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGA 5432
                          |||   |||                 :::
2571 .....................TyrLysLys..............MetTh 2575

5433 GAGGACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCG 5482
      ::::::   |||   :::      ::::::::::::::::::
2575 rAsnAlaMetArgGluPheTyrTyrSerGlnLysValThrLeuTyrGluL 2592

5483 CCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCG... 5529
       :::|||   |||         ::::::   ::::::::::|||
2592 euHisGlyLysLeu.........SerSerTyrTrpAspGluLeuLysThr 2605

5530 ...GGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAAC 5576
        ||||||          :::      ||||||
2606 ValGlyPheAspArgLysLeuAlaLysMetAspLeu.............. 2617

5577 ACAAGACGGCTTGCGTGTGAGAGACCTACCGTAC...GTACTCATAGCTG 5623
      :::|||   :::   |||   |||:::         |||::::::::::
2618 .AspAspAspValSerValValAspPheAsnLeuLysValIleValGlyA 2634

5624 AAAAAGGTATC......TTTACCAAGGGCAAAGATGTCGACGCGGTGGTA 5667
     ::   |||         |||   :::      :::          :::
2634 rgHisGlyThrArgProPheArgAsnThrValAsnSerHisGluPheMet 2650

5668 GCTTTGGGCGACAATCTGTTCGTATGCGAC..............GATAT 5702
               |||:::|||   :::|||                  |||::
2651 PheCysAspAspGlyLeuValLeuCysProGlyProLysLysArgAspPh 2667

5703 ACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTC 5752
     :   ::::::   :::      ::::::::::::|||   :::      |
2667 eAlaLeuIleSerSerGlnThrGlnPheValAlaAlaAsnSerPheLeuA 2684

5753 GATGCGGCATGGTGGGCGAA......TCGTTTAAGTCCTTCGAATATAAG 5796
     ||   ::::::   |||:::  .   :::|||   :::   |||::::::
2684 rg...AlaValAspGlyLysAspLeuThrPheThrAsnSerGluHisSer 2699

5797 ......TGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGT 5840
           |||:::|||||||||||||||||||||||||   :::||||||
2700 LeuIleValTyrGluAlaProProGlyGlyGlyLysThrHisSerLeuVa 2716

5841 GGACGAATTCGTTAAG......TCACCCAATAGCACAGCCACCATTACGG 5884
     |      |||:::          :::      |||      ::::::||||
2716 lLysTyrPheAlaAspTyrCysThrLysValSerCysLeuValValThrA 2733

5885 CTAATGTGGGAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGAT 5934
     |||||   :::|||:::  .:::|||:::::   ::::::::::::
2733 laAsnLysAsnSerGlnAlaGluIleSerGlnArgIleSerArgGluLeu 2749

5935 CCGAATTTGGAAGGTCTC.......................AACAGTGC 5960
        :::   :::   |||                      :::   ::
2750 MetAspArgLysMetLeuAlaLysHisValValLysThrAlaGlyArgVa 2766

5961 TACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATA 6010
     :   ||||||:::|||   ::::::|||   |||   |||
2766 lPheThrValAspSerTyrLeuMetAsnHisIle...ArgLeuLysThrA 2782

6011 AAAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAA 6060
          ::::::|||||||||      |||:::|||   |||   :::
2782 spLeuLeuPheValAspGluCysPheMetValHisAlaGlyAlaIleGly 2798

6061 CTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACAT 6110
        |||    |||:::::::::       :::||||||||||||
2799 AlaValValGluPheThrSerCysLysAlaValValPhePheGlyAspSe 2815
```

```
    6111 AAATCAGATACCATTCATAAACAGGGAGAAGGTG.........TTTAGGA 6151
          :::||||||   :::|||:::|||:::              :::
    2815 rArgGlnIleHisTyrIleHisArgAsnAspPheGlyValSerLeuLeuH 2832

    6152 TGGATTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCT 6201
           |||      :::|||:::   :::   :::      :::|||   :::
    2832 isAspIleAspAlaPheIle...GluProGlnHisArgIleTyrGlyGlu 2847

    6202 AAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGAC 6251
           ||||||||||||||   |||:::|||      ||||||
    2848 ValSerTyrArgCysProTrpAspIleCysGluTrpLeuSer........ 2861

    6252 CGTAAGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGG 6301
                         |||   ||||||:::   |||:::::::::::
    2862 ................LysPheTyrProArgHisValAlaThrAlaAsnT 2873

    6302 ACAAACCAGTAGTAAGA...TCGCTGTCCAAAAGGCCAATTGGAACCACT 6348
                 :::   :::   |||:::|||   :::   |||:::
    2873 hrGlySerIleGlyLysSerSerValSerIleGluIleIleAsnGlyCys 2889

    6349 GATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTT 6398
         |||||||||          ::::::      |||:::      ||||||
    2890 AspAspValProTyrAspSerSerAlaLysTyrIleValTyrThrGlnAl 2906

    6399 GGAGAAGTCGGATATGAAGAGGTCGTTG............AAGGGAAAAG 6436
          ||||||:::::::::::::::   |||               |||:::
    2906 aGluLysAsnGluLeuGlnLysHisLeuGlyArgLeuThrValGlyArgT 2923

    6437 GAAAAGAAACACCAGTGATG...ACAGTGCATGAAGCACAGGGAAAAACA 6483
           |||   :::|||::::::   |||||||||||:::||||||:::|||
    2923 hrLysAlaValProIleValAsnThrValHisGluValGlnGlyGluThr 2939

    6484 TTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATT 6533
         ::::::   |||   |||   |||   |||      :::|||:::   ||
    2940 TyrLysLysValArgLeuValArgCysLysTyrGlnGluAspThrProPh 2956

    6534 CACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCAC 6583
         |   ::::::   |||:::::|||:::|||:::|||||||:::   ||||
    2956 eCysSerAspAsnHisValValValAlaLeuThrArgHisValAspSerL 2973

    6584 TGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATAT 6633
         ||:::|||::::::|||:::|||:::      ||||||:::::::::
    2973 euThrTyrSerValLeuAsnSerArgArgTyrAspLysThrAlaSerSer 2989

    6634 ATTAGCGACGCG 6645
         |||::::::|||
    2990 IleAspGluAla 2993

seq_name: sp_virus:P87587    .

seq_documentation_block:
ID   P87587      PRELIMINARY;      PRT;  3107 AA.
AC   P87587;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P346.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID-12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-VT;
RX   MEDLINE-96406950; PubMed-8811037;
RA   Mawassi M., Mietkiewska E., Gofman R., Yang G., Bar-Joseph M.;
RT   "Unusual sequence relationships between two isolates of citrus
RT   tristeza virus.";
RL   J. Gen. Virol. 77:2359-2364(1996).
DR   EMBL; U56902; AAB38755.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   3107 AA;  346659 MW;  02B748F693BE89CB CRC64;


alignment_scores:
          Quality:  668.00            Length:   2372
            Ratio:  0.615               Gaps:    106
 Percent Similarity:  45.826   Percent Identity: 20.911

alignment_block:
 US-09-301-906-3 x P87587   ..

 Align seg 1/1  to: P87587  from: 1  to: 3107

     892 ATGCGCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAA 941
         :::|||:::   |||   ||||||   :::   |||      :::|||::
     885 IleArgHisIleAlaGluValCysLeuSerMetGlyLysValPhePheAr 901

     942 CAATGAGAGGCTGAGGGGA.............................. 960
         :::::::         |||
     901 gArgAspValAsnLeuGlyAlaPheProTyrValPheGluValGlnHisA 918

     961 ........GAAATTTACAGACGTCGTTGCTTCTCATCTTCCTTTTCGATA 1002
                 :::::::|||   :::|||:::
     918 rgLeuGluArgLeuTyrGlyLysArgAla..................... 927

    1003 GGTTTCTTGATGCACCTTGGCTTTAGATCGTTAAAGGTCATTAGGTTTGC 1052
                 ::::::   |||   |||:::              |||::
     928 .........ValGluTyrGlyValArgGlyTyr............PheSe 937

    1053 GGGCACGAACATACTACAC.........................ATGC 1075
         ::::|||::::::   |||                         :::
     937 rAlaThrArgLeuTyrHisCysCysPheThrAspSerProArgProPheL 954

    1076 CATCACTCAATGAAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGTCTAT 1125
           |||||||||:::   :::   :::|||   :::   |||   :::
     954 ysSerLeuAsnLysTyrHisLysMetGlyGlyGluGluGlyAlaLeuLeu 970

    1126 CTCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACACGGTT 1175
         :::   ::::::                          ||||||   |||
     971 IleThrAspThr.....................AspArgLeuArgAl 979

    1176 GGGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGG 1225
             |||:::         :::|||   :::   |||   :::
     979 aValGlySerAlaTyrGluLysValLysGlyArgLeuProAsp....... 993

    1226 TCTATTCATCGGTCGTTTCGAGTATCACCAATAGACTGGTATTAAGGGAC 1275
               |||   :::|||   |||:::   :::   |||:::       :::
     994 .....SerLeuLeuValArgSerValGluLysAspLeuIleGluPheAsn 1008

    1276 CAATCGGCATTGCTTTCCCATTTGGACACGAAATTGTGCGATATGTTTTC 1325
         |||:::               :::||||||:::
    1009 GlnThrLeu............ValAspThrGln................. 1015

    1326 TCAAAGGGACGCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTGT 1375
                           |||   :::|||:::
    1016 ..................ArgAlaArgProGlyValAlaValProPheG 1026

    1376 TTCTGAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTT 1425
           ::::::   :::::::::      |||   :::   :::||||||:::
    1026 lnMetSerGluAsnGlnGlnValMetLeuThrArgAlaTyrProGluPhe 1042

    1426 TCGATACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGC 1475
         :::|||:::|||:::   ||||||:::|||:::||||||   |||   ::
    1043 AsnIleAsnPheThrHisSerValHisSerAspHisProValAlaAlaGl 1059

    1476 CATGCGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGT 1525
         :   |||         |||   :::      :::   :::|||:::
    1059 ySerArgMetLeuGluAsnHisLeuValArgLysTyrAlaGlySer...A 1075

    1526 GCTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCAT 1575
           :::   |||:::||||||         :::|||:::::|||||||||
    1075 spTyrSerAspValGlyGlyCysProLeuPheHisLeuArgAlaGlyHis 1091
```

```
1576 GTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAGCG 1625
     ::::::   ||||||||||:::||||||   |||:::||||||:::   ||
1092 ThrGlyValHisValCysArgProValTyrAspIleLysAspAlaHisAr 1108

1626 GAGAATC.................AATGAGATCCTCTTTCTTTCCACAG 1657
     ||||:::                :::::::::   |||      |||
1108 gArgValValArgHisHisGlnLeuSerLysValAlaPheAspHisThrA 1125

1658 CTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCGGCTTCA 1707
       |||:::   :::                              :::|||
1125 spGlySerLysGlnVal.......................GlyAlaVal 1133

1708 AAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATTCTAGAGC 1757
     ::::::   :::   |||:::::::          :::|||         ::
1134 AsnThrAsnThrValCysGlyAsnValLeuGlyGluCysPheHisAlaSe 1150

1758 CGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAGCAGGTAG 1807
     :||||||      ||||||:::||||||||:::      ::::::::::
1150 rAspAlaIleValMetValGlnValTyrAspValProLeuArgGluLeuC 1167

1808 CAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGCTCTTATG 1857
     :::::||||||   :::|||   |||   |||      ::::::::::::
1167 ysArgAlaMetValAsnLysLysAlaSerValCysTyrMetThrValVal 1183

1858 TTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGGAAGAACT 1907
       |||   |||||||||   :::      :::   ::::::::::|||||
1184 ThrProGlyGluLeuLeuAspAlaCysGluSerPhePheIleLysGluLe 1200

1908 CGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAATGTTGGTC 1957
     ||||      |||      |||   |||            :::|||
1200 uAspCysSerVal......GluLeuAspArgSerProAsnArgValIleT 1215

1958 AGTGT......GGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGG 2001
       |||      ::::::   |||   |||::::::|||::::::
1215 yrCysPheAsnAsnSerArgTyrThrHisThrTyrSerThrIleCys... 1230

2002 TTTTTCACCTTTTCTTATGTACGCACTTCG............TCCGGGAA 2039
                :::|||:::||||||                :::|||
1231 ............GluTyrMetArgThrProCysValValValAspGlyPh 1243

2040 CGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCA 2089
       :::|||   :::|||      :::   ||||||      ::::::
1243 eLeuPheThrIleGluMetValSerLeuArgCysArgValAsnTyrTyrC 1260

2090 CTATGTGTAGGGCTCAGAAGTCACCTGGA............ACTGAGGTT 2127
       :::   :::::::::      |||                 ::::::
1260 ysIleThrLysSerArgValCysProArgIleSerGluIleLysArgLeu 1276

2128 ACGTATAGGTCGTTG.......................GTCCCGTCGTT 2153
       ||||||                             :::|||   ::
1277 ArgTyrArgArgCysAspSerAspLeuIleArgIleLysIleProArgTy 1293

2154 CGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTAGTGTGT 2203
     :   :::|||:::         :::|||                    :
1293 rSerAsnLysThrArgSerCysLeuPro.....................G 1303

2204 CCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGGATCTATTCC 2253
     ::        :::   ||||||:::   ||||||:::|||:::|||:::
1303 lyCysTyrTyrLeuTyrLeuAspAlaLysPheValSerArgValTyrGlu 1319

2254 TACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTTTGAGTATGC 2303
     |||:::::::|||:::      :::   :::   :::|||||||||:::::
1320 TyrValValAsnAsnCysValValValAsnAlaLysThrPheGluTrpTh 1336

2304 CGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAGGGAGTCGCG 2353
     :   :::   ::::::|||   |||::::::|||:::   ::::::::::
1336 rTrpAsnTyrIleLysSerCysLysSerArgValValIleSerGlyLysI 1353

2354 TTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTT 2403
     :::::|||      |||   |||:::   :::   :::   |||:::
1353 leIleHisArgAspValProIleGlyLeuGluTyrLeuAspGlyPheSer 1369
```

```
2404 GTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAGTATTCGCTC 2453
     ::::::::::::::::   ::::::|||   |||   ::::::
1370 AlaValMetLeuSerAlaGlyValLysGlyArgGlnAsnAlaGluThrPh 1386

2454 CTATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGC 2503
        :::   :::::::|||   :::|||:::      ::::::|||:::|
1386 eSerArgArgLeuArgAlaPheSerGlyAsp...ProSerLeuPheGluL 1402

2504 TC............TTCTTTCAGACCGTAGGCGATTGTTTT......... 2532
     ||           :::   :::         |||   |||
1402 euValArgPheAlaIleSerGluLysCysArgAspLeuPheValGlyVal 1418

2533 TCGAACGCAGTCTCCGTCTATGCTAAGGCAATGGTGCACGATAACTTCAA 2582
     ::::::||||||:::   |||:::|||:::   ::::::::::::|||||
1419 GlnGluAlaValGluArgTyrValLysSerTyrLeuArgAsnSerPheAs 1435

2583 CGTT..................................................... 2586
     |:::
1435 nIleSerPheAlaAspLeuSerAspProLeuLeuThrIleSerGluTyrS 1452

2587 .....TTGGAGACGCTTATGTCTATGCCC..................... 2610
          |||::::::   :::::::::|||
1452 erGluLeuAspValProIleAspLeuProGlyPheGlyCysIleThrGlu 1468

2611 ................................AGAGCGTTCATCCG 2624
                                    |||:::::::||
1469 GlyAspGluAlaLysLeuIleGluAspGlyValLysSerAlaLeuLeuAr 1485

2625 TAAAGTA..............................CCTGGGTCTG 2641
     ||||:::                                 |||:::
1485 gLysAlaPheValThrArgProGlnLeuIleLeuGluArgGlyGlyThrS 1502

2642 TTGTTGTTACCATTTGC.................ACTTCTGGAGCTTCA 2673
          ::::::   |||                 |||   :::|||
1502 erGluIleValProCysGlyAspAspAspGlyLeuGlySerAspGlySer 1518

2674 GACAGGTTG.......................................... 2682
     |||   |||
1519 AspSerLeuLeuGlySerGlySerAspSerSerSerLysAlaSerProGl 1535

2683 ..............................................GAGC 2686
                                                   :::|
1535 yGlyLeuArgAlaGlyArgProSerLeuPheArgSerLeuLeuLysLysL 1552

2687 TCAGGGGTGCCTTTGATATTTCGAAGGAG..................... 2715
     ||   |||:::|||   ::::::|||:::
1552 euGlyGlyGlyPheCysMetAlaLysArgLeuLeuAspLeuLeuCysGlu 1568

2716 ............ACCTTCGGTAGG...AAACTGAAGAATAGTCGCTTGCG 2750
                 ::::::::::|||   |||:::   ::::::::::|||
1569 ArgLeuValGlySerLeuSerArgPheLysPheGlyThrThrHisLeuTh 1585

2751 CGTCTTC.....................TCTAGGGCTATCGTGGAAG 2776
        |||                      ||||||   :::   :::|
1585 rSerPhePheSerMetLeuHisAlaLysLeuSerArgAsnPhePheAspA 1602

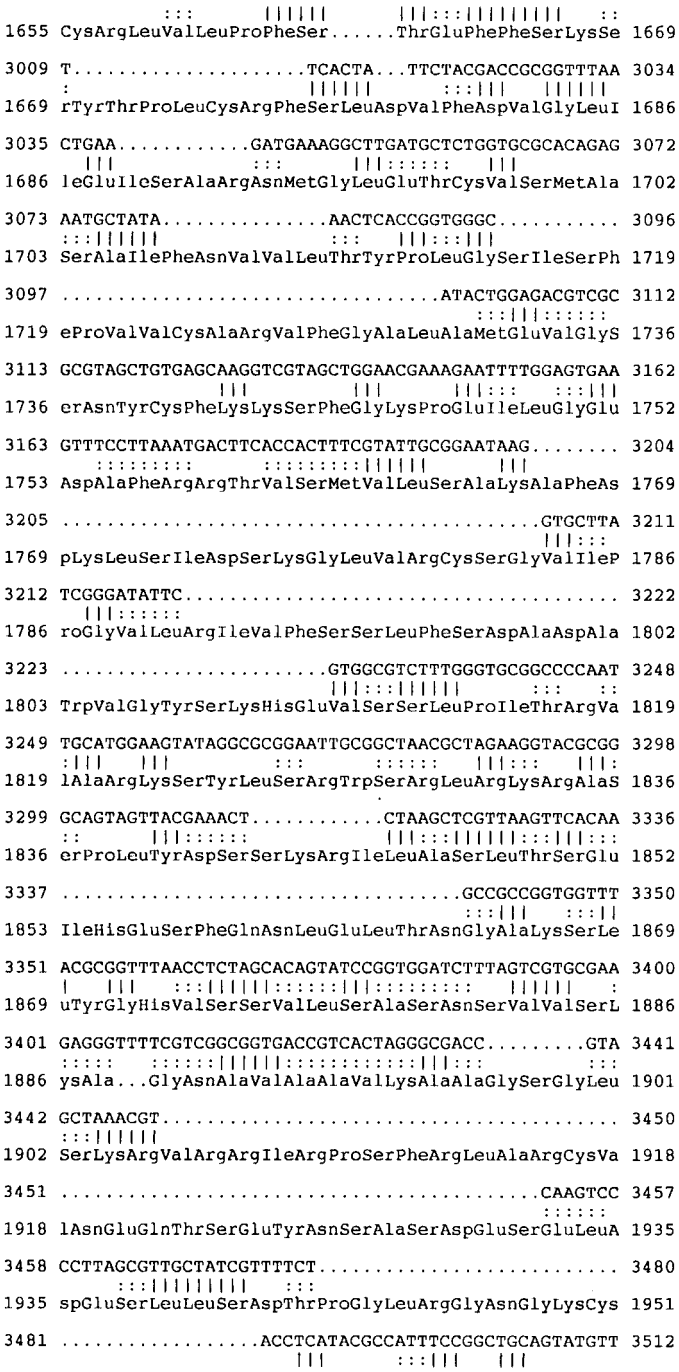
2777 ATTCAATT.................AAGGTCATGAAGGCAATGAAGACA 2808
     ||||||||                |||||:::   ::::::
1602 spSerIlePheThrAlaValTrpAspValValMetAsnLeuIleArgIle 1618

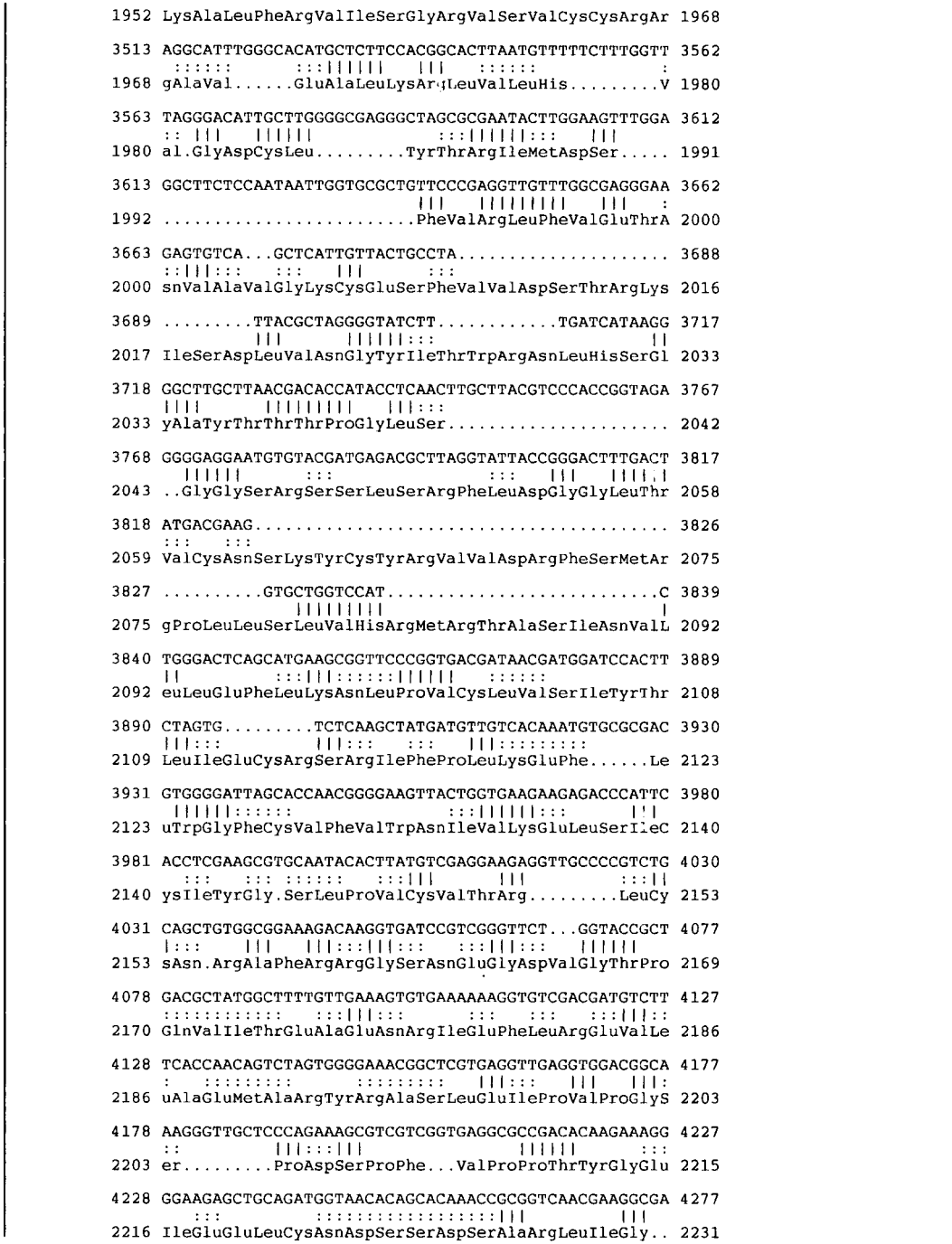
2809 GAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCAT 2858
        ::::||            :::|||:::   ::::::   :::|||
1619 ValAsnGly............ValThrArgThrLysLeuGluCysPhe.. 1630

2859 AATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTGGCG 2908
                   ||||||                   |||   :::|
1631 .............ArgAsnValLeuArgLeuIleProThrLeuCysSerG 1643

2909 GTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGT 2958
     ||   |||            :::   :::|||   ::::::
1643 lyCysLys..............AspPheAlaLysAspPheAlaAsnMet 1654

2959 GGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGG 3008
```


1655 CysArgLeuValLeuProPheSer.....ThrGluPhePheSerLysSe 1669
1669 rTyrThrProLeuCysArgPheSerLeuAspValPheAspValGlyLeuI 1686
1686 leGluIleSerAlaArgAsnMetGlyLeuGluThrCysValSerMetAla 1702
1703 SerAlaIlePheAsnValValLeuThrTyrProLeuGlySerIleSerPh 1719
1719 eProValValCysAlaArgValPheGlyAlaLeuAlaMetGluValGlyS 1736
1736 erAsnTyrCysPheLysLysSerPheGlyLysProGluIleLeuGlyGlu 1752
1753 AspAlaPheArgArgThrValSerMetValLeuSerAlaLysAlaPheAs 1769
1769 pLysLeuSerIleAspSerLysGlyLeuValArgCysSerGlyValIleP 1786
1786 roGlyValLeuArgIleValPheSerSerLeuPheSerAspAlaAspAla 1802
1803 TrpValGlyTyrSerLysHisGluValSerSerLeuProIleThrArgVa 1819
1819 lAlaArgLysSerTyrLeuSerArgTrpSerArgLeuArgLysArgAlaS 1836
1836 erProLeuTyrAspSerSerLysArgIleLeuAlaSerLeuThrSerGlu 1852
1853 IleHisGluSerPheGlnAsnLeuGluLeuThrAsnGlyAlaLysSerLe 1869
1869 uTyrGlyHisValSerSerValLeuSerAlaSerAsnSerValValSerL 1886
1886 ysAla...GlyAsnAlaValAlaAlaValLysAlaAlaGlySerGlyLeu 1901
1902 SerLysArgValArgArgIleArgProSerPheArgLeuAlaArgCysVa 1918
1918 lAsnGluGlnThrSerGluTyrAsnSerAlaSerAspGluSerGluLeuA 1935
1935 spGluSerLeuLeuSerAspThrProGlyLeuArgGlyAsnGlyLysCys 1951


1952 LysAlaLeuPheArgValIleSerGlyArgValSerValCysCysArgAr 1968
1968 gAlaVal.....GluAlaLeuLysArgLeuValLeuHis.....V 1980
1980 al.GlyAspCysLeu.....TyrThrArgIleMetAspSer..... 1991
1992 PheValArgLeuPheValGluThrA 2000
2000 snValAlaValGlyLysCysGluSerPheValValAspSerThrArgLys 2016
2017 IleSerAspLeuValAsnGlyTyrIleThrTrpArgAsnLeuHisSerGl 2033
2033 yAlaTyrThrThrThrProGlyLeuSer..... 2042
2043 ..GlyGlySerArgSerSerLeuSerArgPheLeuAspGlyGlyLeuThr 2058
2059 ValCysAsnSerLysTyrCysTyrArgValValAspArgPheSerMetAr 2075
2075 gProLeuLeuSerLeuValHisArgMetArgThrAlaSerIleAsnValL 2092
2092 euLeuGluPheLeuLysAsnLeuProValCysLeuValSerIleTyrThr 2108
2109 LeuIleGluCysArgSerArgIleProProLeuLysGluPhe.....Le 2123
2123 uTrpGlyPheCysValPheValTrpAsnIleValLysGluLeuSerIleC 2140
2140 ysIleTyrGly.SerLeuProValCysValThrArg.....LeuCy 2153
2153 sAsn.ArgAlaPheArgArgGlySerAsnGluGlyAspValGlyThrPro 2169
2170 GlnValIleThrGluAlaGluAsnArgIleGluPheLeuArgGluValLe 2186
2186 uAlaGluMetAlaArgTyrArgAlaSerLeuGluIleProValProGlyS 2203
2203 er.....ProAspSerProPhe...ValProProThrTyrGlyGlu 2215
2216 IleGluGluLeuCysAsnAspSerSerAspSerAlaArgLeuIleGly.. 2231

```
4278 CAGGGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATTC 4327
                ||||||||||||||||   |||                 :::|
2232 ..........ValAspAsnGlyLeuAspSer..............ValP 2240

4328 CAAAGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAA 4377
     ||          ::::::|||:::   ::::::   :::       ::::::
2240 roThrGluGlyGluAlaGluLeuAspSerAspGlyAspSerAspAspSer 2256

4378 GTACCATTG.....................GCGACTGTTTCGGGCGCCAC 4406
     :::||||||                     :::   ::::::|||:::
2257 IleProLeuAspArgAspPheGluGlnIleGlyGlyLeuGlyGlySerGl 2273

4407 GCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGA 4456
         :::                     :::|||         |||:::|
2273 yAsnVal.....................AspValPheArgLeuGlyLeuL 2283

4457 AGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAG 4506
     ||::::::         :::   :::   ::::::   :::   :::
2283 ysPheVal......TrpArgMetLeuArgAsnLeuPheAsnLeuArgAla 2297

4507 GTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACGATCAATGAAGGCAA 4556
     |||:::   :::                 :::      :::|||   |||::
2298 ValLysPheSerTyrGlyLeuAlaLeuGlnCysIlePheAsnGlyGlySe 2314

4557 GGCCGGT..................................... 4563
     :  |||
2314 rMetGlyGlnPheThrSerArgLeuLeuPheSerPheGlyAsnCysLeuP 2331

4564 .............AAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGT 4599
                 :::::::::|||      |||      |||   :::
2331 roPheGluLeuPheArgHisValCysSerValArgTrpLeuSerLysSer 2347

4600 GTGCAGCTGGATGTGTACAAC.............GAAGCGACTATCGC 4634
     :::   :::   |||   |||               |||   |||:::
2348 LeuAlaIleArgValArgAsnAlaArgIleCysAspGluGluThrValGl 2364

4635 CACCAGGTTC.................................TCAA 4648
          |||                                    |||:
2364 uArgPhePheSerPheSerArgIleLeuTyrAspSerArgTyrLeuSerA 2381

4649 ACGCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTC 4698
     :::::|||:::   ::::::::::::           :::   :::|||
2381 spValPheSerValLeuGluArgValAspPheTyrLeuSerTyrLeuPhe 2397

4699 TCAAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATC 4748
     |||         :::      |||   |||   :::   :::|||:::
2398 SerLeuProLysArgCysLeuThrArgAsnArgSerValArgValThrAr 2414

4749 AGGGTGGCCT...CGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACC 4795
          |||   |||::::::::::|||:::
2414 gArgAlaProValArgValIleArgAspMet..................  2424

4796 CAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGC 4845
                    ||||||:::      |||   :::
2425 ................GlyLeuValGluAspGlyLysLysSer......  2433

4846 GTACCATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTAT 4895

2433 .....................................................  2433

4896 CTTGACGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGA 4945
      ||||||     |||   :::      ::::::::::::::   ||||
2434 .LeuThrGluValLeuAsnSerLeuAspSerLeuLysValSerAspArgL 2450

4946 AGGGTGGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTT 4995
     |||||   ::::::::::::           :::::|||
2450 ysGlyLysSerAlaLeuIle............GluSerAsp......... 2459

4996 ATGCCTTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGA 5045
          |||   :::             ||||||   :::|||||||::
2460 .........GlySerAspSerGluSerGlyHisSerPheGlySerIleSe 2473

5046 CGAAGGGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTG 5095
     :   |||   ::::::
2473 rPheGlySerValLys..................................  2478

5096 TAGGCAGGATGTTGCAG...TTAGCCGGCGGCGTGTCGGATGAGAAGTCA 5142
          ::::::|||:::   :::   :::|||||||   :::|||||||:::
2479 ..GluGluValLeuGluAlaValAsnSerGlyValValSerGluLysAla 2494

5143 CCAGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAAT 5192
     |||   :::   :::   :::       |||:::                ::
2495 ProLeuMetLysHisAlaAspAlaLeuSerArg..............Le 2506

5193 CACACCAAAATCGGGGGGCAAG..................................  5214
     :   |||       :::||||||
2506 uGluProAspLeuSerGlyLysLysAspLeuProIleValArgArgSerP 2523

5215 ..GCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATAC 5262
       :::   :::|||::::::|||   :::       |||
2523 heThrArgGlyGluSerThrGlyThrLysProAlaGly............ 2535

5263 TCGATATGGTGCGAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGC 5312
                                        |||   :::|||   ::
2536 ..........................SerLeuCysValTyrLeuThrTh 2543

5313 TGATAATCCAGTGATGGCTCTTGAACCTGACTACACCCCAATGACATTTG 5362
     :   |||       ::::::       ||||||      |||::::::  :
2543 rLeuAsnSerSerPheGlySerValProAspLeuTyrProValAlaArgA 2560

5363 AAGTGGTTAAAACCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAG 5412
     :::::
2560 spIle.....................................  2561

5413 TATCTGGCTATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCT.. 5460
                                  ||||||:::   :::       |||
2562 .........................ThrTyrLysLysMetThrAsnAlaMe 2570

5461 ...................AGAAATATTGCCGTCACTACCGCCGAAGGTG 5491
                         :::::::::::::::          :::|||
2570 tProLysPheTyrTyrSerGlnLysValThrLeuTyrGluLeuHisGlyL 2587

5492 TTCTGAAAGTACCTAATCAAGTTTATGAATCACTACCG......GGCTTT 5535
       |||         ::::::   ::::::::::|||         ||||||
2587 ysLeu.........SerSerTyrTrpAspGluLeuLysThrValGlyPhe 2600

5536 CACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGG 5585
     :::         :::      ||||||               :::|||
2601 AsnArgLysLeuAlaLysMetAspLeu..............AspAspAs 2612

5586 CTTGCGTGTGAGAGACCTACCGTAC...GTACTCATAGCTGAAAAAGGTA 5632
       :::   |||   |||:::          |||:::::::::::::   |||
2612 pValAspValValAspPheAsnLeuLysValIleValGlyArgHisGlyT 2629

5633 TC......TTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCTTTGGGC 5676
           |||   :::      :::            :::
2629 hrArgProPheArgAsnThrValAsnSerHisGluPheMetPheCysAsp 2645

5677 GACAATCTGTTCGTATGCGAC..............GATATACTGGTTTT 5711
     |||:::|||   :::|||                |||:::   :::::
2646 AspGlyLeuValLeuCysProGlyProLysLysArgAspPheAlaLeuIl 2662

5712 CCACGATGCCATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCA 5761
     :   :::      :::::::::::|||   :::      |||   ::::
2662 eSerSerGlnThrGlnPheValAlaAlaAsnSerPheLeuArg...AlaV 2678

5762 TGGTGGGCGAA......TCGTTTAAGTCCTTCGAATATAAG......TGC 5799
     ::   |||:::      :::|||   :::   |||::::::
2678 alAspGlyLysAspLeuThrPheThrAsnSerGluHisSerLeuIleVal 2694

5800 TATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGACGAATT 5849
     |||:::|||   ||||||||||||||||   :::||||||      ||
2695 TyrGluAlaArgProGlyGlyGlyLysThrHisSerLeuValThrTyrPh 2711

5850 CGTTAAG......TCACCCAATAGCACAGCCACCATTACGGCTAATGTGG 5893
```

```
2711 eAlaArgTyrGlyThrLysValSerCysLeuValValThrAlaAsnLysA 2728

5894 GAAGTTCTGAGGACATAAATATGGCGGTGAAGAAGAGAGATCCGAATTTG 5943
2728 snCysGlnAlaGluIleSerGlnArgIleSerArgGluLeuMetAspArg 2744

5944 GAAGGTCTC.....AACAGTGCTACCACAGT 5969
2745 LysMetLeuAlaLysHisValValLysThrAlaGlyArgValPheThrVa 2761

5970 TAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTT 6019
2761 lAspSerTyrLeuMetAsnHisIle...ArgLeuLysThrAspLeuLeuP 2777

6020 TGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTC 6069
2777 heValAspValCysPheMetValHisAlaGlyAlaIleGlyAlaValVal 2793

6070 TTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGAT 6119
2794 GluPheThrSerArgLysAlaValValPhePheGlyAspSerArgGlnIl 2810

6120 ACCATTCATAAACAGGGAGAAGGTG.....TTTAGGATGGATTGTG 6160
2810 eHisTyrIleHisArgAsnAspPheGlyValSerLeuLeuHisAspIleA 2827

6161 CTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGTAC 6210
2827 spAlaPheIle...GluProGlnHisArgIleTyrGlyGluValSerTyr 2842

6211 AGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGGGG 6260
2843 ArgCysProTrpAspIleCysGluTrpLeuSer..... 2853

6261 AACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAG 6310
2854 .....LysPheTyrProArgHisValAlaThrAlaAsnThrGlySerI 2868

6311 TAGTAAGA...TCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTA 6357
2868 leGlyArgSerSerValSerIleGluIleIleAsnGlyCysAspAspVal 2884

6358 GCTGAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTC 6407
2885 ProTyrAspSerSerAlaLysTyrIleValTyrThrGlnAlaGluLysAs 2901

6408 GGATATGAAGAGGTCGTTGAAGGGAAAA.....GGAAAAG 6442
2901 nGluLeuGlnLysHisLeu...GlyArgLeuThrValAlaGlyThrLysA 2917

6443 AAACACCAGTGATG...ACAGTGCATGAAGCACAGGGAAAAACATTCAGT 6489
2917 laValProIleValAsnThrValHisGluValGlnGlyGluThrTyrLys 2933

6490 GATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTCACTAA 6539
2934 LysValArgLeuValArgCysLysTyrGlnGluAspThrProSerCysSe 2950

6540 ACAACCGCATATACTTGTTGGTTTGTCGAGACACACACGCTCACTGGTTT 6589
2950 rAspAsnHisValValValAlaLeuThrArgHisValAspSerLeuThrT 2967

6590 ATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCACATATATTAGC 6639
2967 yrSerValLeuAsnSerArgArgTyrAspLysThrAlaSerSerIleAsp 2983

6640 GACGCGTCACCTCAATCAGTATCCGACGCTTTGCTTCACACGTTCGCCCC 6689
2984 GluAla......ArgGluIlePheAspLysPheArgSerThrAsnHisSe 2998

6690 GGCTGGTTGC 6699
```

2998 rHisGlyCys 3001

seq_name: sp_virus:Q83044

seq_documentation_block:

```
ID   Q83044      PRELIMINARY;      PRT;  1873 AA.
AC   Q83044;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   METHYLTRANSFERASE.
OS   Lettuce infectious yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=31713;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92;
RA   Klaassen V.A., Boeshore M., Koonin E.V., Tongyan T., Falk B.W.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U15440; AAA61797.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   1873 AA;  217264 MW;  55A771D8357761F0 CRC64;
```

alignment_scores:

```
            Quality:  652.50            Length:   2227
              Ratio:  0.690               Gaps:     84
 Percent Similarity:  42.434   Percent Identity: 19.129
```

alignment_block:

US-09-301-906-3 x Q83044 ..

Align seg 1/1 to: Q83044 from: 1 to: 1873

```
 496 GTTAAGCCAACAGCTGTTGGCGTCCACGTACCACTTCCTAAAAAACAGGA 545
259 ValLysProArgValValSerValAsn.....LysAlaAr 270

 546 AGCACTGGAGCCAGCGCAATCAGTCCCACAACAGTCGTTGGAGGAGAAGG 595
270 gGluValSerSerServalGluValLysLeuGlnArgTyrGluArgGluG 287

 596 CCGCCTTGACGTTTGGCCTTTTCTTCAGTAAAGGTGGGGGTGATGAGAGC 645
287 lyGluVal..... 289

 646 GACGCTGTCATCTTGCGGAAAGGGAAATTGTTTAACAGGGCCCTTAATGT 695
290 ...PheValValSerLysAlaGlyLysGluAspThrArgValLeuAsnAs 305

 696 TCCTATTGATGTAAAGAACACGTTCGTTTGGGCTAAAATCTGGGATGAAG 745
305 pAspAsnAlaIleArgAsnLeuPhe.....AsnA 315

 746 CCTCTCGTAGGAGAGGGTATTTTTACGTCAAAGATAGAGCTGTTAAATTC 795
315 laThrLeuAsnGlyGlyLysTyr.....LysLeu 324

 796 TTCCCTATTGTGCGGGGTAGGGCTACGATCGAGGACTTCATCGTGAATAC 845
325 HisPro..... 326

 846 AGCCCCAGGGTGTGATGTTGCCTTGCCGCGCATTGAGTTGTGGAGTATGC 895
327 .....GluAlaLysSerAlaT 332

 896 GCGAAAGGGCGTTTGTATGCACCACCAAAGGGTGGTGTTGGTTTAAC... 942
```

```
 332 hrGlyLysArgPheLysTyrTyrLysAspGlyPheCysTrpLeuAspVal 348

 943 .....AATGAGAGGCTGAGGGGAGAAATTTACAGACGTCGTTG 980
               |||:::|||:::     |||      :::   :::||
 349 PheAlaAspAlaAsnArgArgIle...ProGluTrpValLysProHisCy 364

 981 CTTCTCATCTTCCTTTTCGATAGGTTTCTTGATGCACCTTGGC..... 1023
     |:::   ::::::         :::   ||||||     |||
 364 sLeuLeuThrGly.........SerValLeuMetSerCysGlyLeuTrpA 378

1024 ..TTTAGATCGTTAAAGGTCATTAGGTTTGCGGGCACGAACATACTACAC 1071
       |||   :::   |||::::::      ::::::   ::::::||||||
 378 spPheAlaLysArgLysMetValSer...ValSerHisGlyLeuLeuHis 393

1072 ATGCCATCACTCAATGAAGAGCGTACCTTTGGGTGGAAGGGCGGAGACGT 1121

 393 ..................................................  393

1122 CTATCTCCCCAATGTCCCAAAAACCGCTATCGTCGCTGGCGATAGGACAC 1171
      |||      ::::::   :::::::::      ||||||   |||
 394 .TyrAspArgLysLeuGluArgSerSerAlaArgAlaGlyValArgAspP 410

1172 GGTTGGGAGGGGAGATCTTGGCCTCCGTCGCCAATGCCCTTAATCAAGAG 1221
       :::|||                ::::::|||         |||:::
 410 heValGly.................AlaSerAsnGluAlaValGlnArg 420

1222 GAGGTCTATTCATCGGTC.........GTTTCGAGTATCACCAATAGACT 1262
     |||   :::   ::::::         |||:::::::::::::::|||::
 421 GluAspPheArgAspLeuAspLeuArgValGluGluPheAlaGluArgVa 437

1263 GGTA............TTAAGGGACCAATCGGCATTGCTTTCCCATTTGG 1300
     ::::            ||||||:::::::::   |||:::::::::::
 437 lLeuGluThrAlaAsnLeuArgSerAspAsnArgLeuIleAspAsnIleL 454

1301 ACACGAAATTGTGCGATATGTTTTCTCAAAGGGACGCAATGATTCGCGAA 1350
       |||:::      |||   ::::::::::::::         :::|||
 454 euThrArgAlaSerAspTyrIleAsnLysLysSerLysGluSerLysGlu 470

1351 AAACCCTCACATAGGTGCGATGTGTTTCTGAAGCCGCGGGAAAGGGAGAA 1400
            :::   |||         |||:::      |||::::::
 471 LeuAspIleAsnValCys.........LeuSerMetAspGluLysLysMe 484

1401 GCTGAGGGAACTCTTTCCAGAGCTTTCGATACAGTTCTCCGACTCGGTCA 1450
      :::   :::||||||||||:::::::::::::::|||:::::::::
 484 tIleThrAsnLeuProAspIleGlnMetSerPheAsnGlnLysSerT 501

1451 GGAGTAGTCACCCATTCGCTAATGCCATGCGGAGCTGTTTCAATGGAATC 1500
       |||:::|||         ||||||||||||:::|||   |||
 501 yrSerAsnHisGlyValPheAsnAlaMetArgAlaCysGluAsnPheTyr 517

1501 TTTTCCAGGAGGTGTGGTAATGTGTGCTTCTTCGATATTGGGGGGAGCTT 1550
     |||||||||:::      |||      :::::|||   ||||||:::
 518 PheSerArgLysPheLysAsnSerAspTyrIleAspAlaGlyGlyAspVa 534

1551 CACGTATCATGTCAAAGCTGGCCATGTGAACTGTCATGTATGCAATCCAG 1600
      :::      :::::::::   :::   |||   |||:::|||:::|||
 534 lValSerThrLeuArgSerLysAsnHisAsnValHisIleCysSerProA 551

1601 TCCTAGACGTTAAAGATGTGAAGCGGAGAATCAATGAGATCCTCTTTCTT 1650
       ||||||:::||||||:::   |||:::|||::::::         :::
 551 rgLeuAspLeuLysAspAlaAlaArgHisIleGlnArgAlaThrValIle 567

1651 TCCACAGCTGGGGGAGATTCGTACGTGTCCAGTGACCTTCTAACTGAAGC 1700
     :::         |||
 568 AspGlyLeuLysGly.............................. 572

1701 GGCTTCAAAGTCTGTGTCTTACTGTAGTCGAGAATCGCAGAACTGCGATT 1750
        ::::::::::::|||:::|||::::::::::::::::::|||
 573 .TyrGlyGluThrIleSerPheCysThrAsnLysThrGluAspCysAlaV 589

1751 CTAGAGCCGATGCGGGTTTTATGGTGGATGTGTACGATATATCCCCGCAG 1800
       :::   |||      :::   |||:::||||||||||::::::   :::
 589 alAsnArgAspIleIleIleAlaValGluValTyrAspMetThrLeuArg 605

1801 CAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGACATAGC 1850
     :::::::|||:::|||||||   :::   |||:::      |||:::::::::
 606 AspMetAlaLysAlaMetLeuSerHisGlySerArgLysPheGluPheSe 622

1851 TCTTATGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTACTTGG 1900
     :   ::::::|||   |||::::::      :::   :::|||   |||
 622 rCysIleIleProProGluIlePheThrLysGluCysAsnValGluLeuT 639

1901 AAGAACTCGATACGTTGGTGAAGAGGGAAGGTGATTACCTGGCCTACAAT 1950
       |||            |||   |||   ||||||   :::   |||
 639 yrGluGlyArgLeuLysValThrArgIleGlyAspAsnValGluTyrTyr 655

1951 GTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGG 2000
       |||:::   ||||||   :::::::|||||   ::::::::::::
 656 TyrGlySerAsnGlyGluThrPheSerHisSerCysGlnThrLeuLysAs 672

2001 GTTTTTCACCTTTTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGC 2050
      :::::::::   :::   |||         :::|||:::|||||||||
 672 pIleLeuSerValGln...ValPheGlnPheGlyGlyArgValPheLysL 688

2051 TAGAGTATGAGGGATACCGTTGTGGTTACCATCATCTCACTATGTGTAGG 2100
           |||      |||         |||   ::::::::::|||
 688 ysThrLeuGluHisSerArgGlyGlnLeuHisPhePheSerIleCysIle 704

2101 GCTCAGAAG...TCACCTGGAACTGAGGTTACGTATAGGTCGTTGGTCCC 2147
     :::::::|||   :::|||||||:::
 705 CysGluLysIleGluProGlySerValLysLeuLysThrTyrTyrGlnAr 721

2148 GTCGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGTAGCTGGTTCTA 2197
      |||   :::   |||      :::   |||||||||      :::   :
 721 gSerGluLeuAspLysValThrLeuArgIleProValLysAspSerPheG 738

2198 GTGTGTCCTTTAAGACAATAGTCCTC...GATTCGGACTTTGTCGACAGG 2244
     ::|||            |||:::      |||   :::||||||:::
 738 lyValValThrHisTyrIleIleLysGluAspArgGluPheValSerSer 754

2245 ATCTATTCCTACGCGCTCAACACTATAGGGACATTCGAGAATAGAACGTT 2294
     :::   :::|||:::   ||||||         ::::::::::::
 755 MetIleGluTyrValAlaAsnThrGlyIleLysIleAspAspLysMetVa 771

2295 TGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTCATTACAG 2344
      |||:::::   :::      |||::::::||||||   :::|||   :
 771 lGluTrpThrTyrSerGlnTyrArgAlaLysLysThrValThrIleLysS 788

2345 GGAGTCGCGTTGTCCACAGCAAGGTTGATATTTCTCCTGATGATATGTGG 2394
     ::::::::|||:::::::::::::::   |||      :::   :::
 788 erGlyLysValThrGlnLysGluThrArgIleArgLysGluLeuIlePro 804

2395 GGTTTAGTTGTCGCTGTTATGGCTCAGGCGATTAAGGATAGGGCGAAGAG 2444
     |||:::::::::   :::|||:::::::::|||:::   |||   |||::
 805 GlyPheIleAlaIleIleMetSerGluGlyIleArgAlaArgGluLysTh 821

2445 TATTCGCTCCTATAACTTTATAAAAGCCAGTGAGGGGAGT..... 2484
     :           :::::::::   :::|||:::   :::
 821 rHisTyrLeuAlaLysMetLeuTyrThrSerHisTyrLysProSerIleV 838

2485 ............................CTCGCCGGGGTCTTCAAG 2502
                                 |||:::|||:::   :::
 838 alAsnIleIlePheArgLeuIleMetHisPheLeuGlyGlyThrLysArg 854

2503 CTCTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTCTA 2552
     ::::::::::::::::::   ||||||:::::      :::
 855 PheIleTyrGluSerLeuValAspCysLeuLysPheLeuThr........ 868

2553 TGCTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGCTTATGT 2602
                          |||   :::   ::::::|||:::::::
 869 ......................AsnSerAspTyrIleAspThrIleValA 878

2603 CTATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACC 2652
     ::
 878 sn..................................................  878
```

```
2653 ATTTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGA 2702
                  :::::::::|||:::|||              ||
 879 ..............ThrGluSerArgIleGlu..............As 885

2703 TATTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCG 2752
     |:::::::||
 885 pLeuAspLys..................................... 888

2753 TCTTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATG 2802
                 :::   |||:::::::::   :::
 889 ........TrpPheValPheGluGlnAsnValThrIle............ 898

2803 AAGACAGAAGATGGAAAACCCCTGCCA...ATTACTGAAGATTCTGTATA 2849
        |||   |||::::::      |||   |||   |||:::||||||
 899 ...ThrThrAspAlaGluAspGlnProSerIleLeuGluGlnSerValLy 914

2850 TGCGTTCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTC 2899
      :::|||:::
 914 sThrPheLeu..................................... 917

2900 TTCTTGGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTG 2949
                                |||:::   |||:::   :::
 918 ..........................SerLysTyrSerGluAspVal 924

2950 GTAGCTCGTGGGGCTGCGACG............AAGGCCTTTTCTGGCAT 2987
           ||||||      :::           :::   ||||||
 925 TyrGluArgGlyAspArgSerGluPheGluLeuGluGluPheSer..... 939

2988 TACGTCGTTCTTTTCCACAGGTTCACTATTCTACGACCGCGGTTTAACTG 3037
                                                  :::|
 940 ..............................................SerG 941

3038 AAGATGAAAGGCTTGATGCTCTGGTGCGCACAGAGAATGCTATAAACTCA 3087
     ||::::::   :::|||   |||                 :::||||||
 941 luGlnAspLeuValAspGlnLeu..................LeuAsnSer 951

3088 CCGGTGGGCATACTGGAGACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGC 3137
                                                    ::
 952 Gly.............................................Gl 953

3138 TGGAACGAAAGAA.........TTTTGGAGTGAAGTTTCCTTAAATGACT 3178
     :|||:::||||||         |||:::|||                 :
 953 yGlySerLysGluAspGluTrpPheTyrSer.................T 964

3179 TCACCACTTTCGTATTGCGGAATAAGGTGCTTATCGGGATATTCGTGGCG 3228
     ::         ::::::|||:::|||      ::::::
 964 yrIlePheSerIleValArgThrLysCysThrLeuSer............ 976

3229 TCTTTGGGTGCGGCCCCAATTGCATGGAAGTATAGGCGCGGAATT..... 3273
                    :::||||||:::::   :::   |||
 977 .................LeuAlaTrpArgTrpThrAsnArgIlePheAs 987

3274 ..............GCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTT 3307
                      |||:::::::::   ::::::         |
 987 nPheIleThrSerCysArgAlaLysGlyLysAspPheValLysTyrIleT 1004

3308 ACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGT 3357
     |||||   ::::::  :::   ::::::                :::::
1004 yrGluIleIleLysIleIleIleAsnLysPro.........TrpGluAla 1017

3358 TTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTT 3407
     |||          |||:::|||   :::   |||::::::
1018 LeuCysHisTyrLeuValGlnGlyTyrLysThrValLeuLys........ 1031

3408 TTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCC 3457
               :::   |||      ::::::      ||||||:::
1032 .IleProMetThrCysValAspSerValAsnTyrLeuLysArgLysSerL 1048

3458 CCTTAGCGTTGCTATCGTTTTC.....................TAC 3482
             :::||||||:::                       |||
1048 euGluLysIleLeuSerTyrPheSerHisGluArgAlaAspAspGluTyr 1064

3483 CTCATACGCCATTTCCGGCTGCAGTATGTTAGGCATTTGGGC........ 3524
```

```
     ::::::      :::      ::::::   ::::::   |||
1065 IlePheProGlyTyrAspSerAspPheSerGluGluSerGlyAspTyrGl 1081

3525 ...............................ACATGCTCTTCCACGGCAC 3543
                                          |||:::
1081 yAspGluAsnThrGluLeuAspSerArgSerArgCysThrValIleLysS 1098

3544 TTAATGTTTTTCTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGC 3593
       :::|||   :::   ||||||:::|||                  :::
1098 erLysValGluMetLeuPheArgAsnIle.................Lys 1108

3594 GAATACTTGGAAGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCC.. 3641
       ::::::|||   :::                |||   ||||||:::
1109 GlyHisIleGluLysPhe..............LeuGlnArgCysGlyIl 1120

3642 ................................CGAGGTTGTTTGGCGA 3657
                                            ::::::   |||:::
1120 eIleAspGlnTyrAsnLysIleTyrHisTrpPheLysSerIleLeuThrA 1137

3658 GGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCTAGGGGTATCTTT 3707
          :::|||         |||      :::                 :::
1137 spValAspCysPheAsnSerValLeuAsnPhe..............Ile 1148

3708 GATCATAAGGGGCTTGCTTAACGACACCATACCTCAACTTGCTTACGTCC 3757
        |||   |||  |||                |||   :::||||||:::
1149 ThrHisThrGly..Cys......AspAlaLeuLeuAlaValLeuThrGly 1162

3758 CACCGGTAGAGGGGAGGAATGTGTACGATGAGACGCTTAGGTATTACCGG 3807
     |||            :::   |||:::::::::
1163 HisPhe...SerIleAlaSerCysAlaLeuLys................. 1172

3808 GACTTTGACTATGACGAAGGTGCTGGTCCATCTGGGACTCAGCATGAAGC 3857
               ||||||   |||:::::::::   :::||||||   |||:
1173 .......PheMetThrAspValValIleGluLysAsnLeuSerGluLysH 1187

3858 GGTTCCCGGTGACGATAACGATGGATCCACTTCTAGTGTCTCAAGCTATG 3907
     ::      :::|||      :::       |||::::::
1187 isHisThrThrThrLysLeuIleAlaGlyAsnLeuThrGlnCysIleTyr 1203

3908 ATGTTGTCA..............CAAATGTGCGCGACGTGGGGATTAGC 3942
        |||:::                  :::         ||||||:::
1204 LysLeuAspPheIleHisProIleTrpIleProIleArgTrpGlyValLy 1220

3943 ACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAGCGT 3992
           :::|||:::   :::||||||:::   ::::::
1220 sGluIleSerLysIleHisIleLysLysLysLeuLeuAsn.......... 1233

3993 GCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGG 4042
                                                  ||||||
1234 ..............................................TrpArg 1235

4043 AAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTT 4092

1235 ................................................... 1235

4093 GTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAG 4142
                           ||||||   :::   ||||||:::   :
1236 .....................SerThrLysGluValThrAsnGluSerI 1245

4143 TGGGGAAACGGCTCGTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAG 4192
     ::   |||   ||||||
1245 leCysLysAspLeuVal.................................. 1250

4193 AAAGCGTCGTCGGTGAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGAT 4242

1250 ................................................... 1250

4243 GGTAACACAGCACAAACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACA 4292
               ||||||                  :::      :::|||
1251 ..........HisLys..................SerTyrIleLysTyrP 1258

4293 GTCCAGTCTTGTGAGTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGT 4342
        :::
```

```
1258 heAspIle..................................... 1260

4343 CCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACT 4392
                                  ||||||     |||
1261 ..........................LysLysAlaValTrp..... 1265

4393 GTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTAC 4442

1265 ............................................... 1265

4443 GACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGG 4492
                                              |||:::|||
1266 ....................................LeuLeuTrp 1268

4493 CTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAGAGGAGTTTGACG 4542

1268 ............................................... 1268

4543 ATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGA..... 4584
                                 |||:::   |||
1269 ........................PheSerLeuLeuIleIlePheL 1276

4585 ........ACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCG 4626
             |||   :::      :::   :::::::::|||
1276 euHisProThrLeuGlyPhePheIleLeuPheSerIleTyrProAlaIle 1292

4627 ACTATCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAA 4676
     |||         ::::::|||:::      |||   :::|||:::||
1293 GluIleLysArgTyrTyrAsnAsnValValCysPhe...AsnAsnIleLy 1308

4677 AGGGAGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTAT. 4725
     |                   :::::::::||||||      |||::::::
1308 sMetSerTyrProHisValIleAspArgLeuGlyGlyThrTyrAsnPheA 1325

4726 .........................................AATGGT 4731
                                               :::
1325 laLysLeuLysLysAlaValArgGluLysPheSerGluAsnLysSerGln 1341

4732 GGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAAC 4781
     |||      :::|||:::         :::      ||||||::::::
1342 GlyValLysIleSerGluIleProAlaGluLysAspGluAspValMetPr 1358

4782 GGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTACA 4831
     ::::::                                |||:::      |
1358 oProLeuGluGlu.....................ValGluThrProL 1367

4832 AGATGGGTGGAGGCGTACCATTCCACGCTGATGAC.........GAGGAG 4872
     ||   :::   |||   |||         :::|||         :::|||
1367 ysArgSerValGlyArgProSerSerLysGlnAspAspIleGluAspGlu 1383

4873 TGCTATCCATCAGATAACCCTATCTTGACGGTCAATCTC........... 4911
     |||      :::::::::|||      :::::::|||:::
1384 CysAsnIleLysGluAspProThrGluSerLeuAsnPheSerAsnIleGl 1400

4912 ............................GTGGGGAAGGCAAACTTC. 4929
                                 ||||||:::::::::::
1400 uAlaThrSerAsnArgTyrAspThrLeuArgValGlyAsnSerLysLeuC 1417

4930 ......................TCGACTAAGTGCAGGAAGGGTGGTAAG 4956
                           ::::::||||||
1417 ysLysPheLeuArgPheTyrProLysSerLysCys.............. 1428

4957 GTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCCTTGCGG 5006
                |||:::   :::||||||   :::
1429 ......TyrAlaAsnIleGlnThrGlyAspLeuIle.............. 1438

5007 TTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAGGGCGCA 5056
           |||      :::::::::         ::::::::::::::::
1439 .......ArgAspSerIleGluGluPheTyrHisLeuGluArgSerLysL 1453

5057 TCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTA......... 5097
     ::::::::   :::         :::||||||      :::
1453 euAspIleGluIleSerLysMetGlnArgValValGluIleMetAsnAsp 1469

5098 ...GGCAGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAGAAGTCACC 5144
        ||||||||||||:::   :::      :::   |||::::::|||
1470 ThrGlyArgMetLeuAspAsnValArgArgMetIleAspAspArgSerMe 1486

5145 AGGTGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCA 5194
        |||
1486 tTyrVal.......................................... 1488

5195 CACCAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTC 5244
                                                   :::
1489 ..........................................ThrLeu 1490

5245 AAGGGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAA 5294
        |||   |||   |||   :::   |||::::::
1491 AspGlyValSerTrpTyrArgLeuGlyCysLysAsp.............. 1502

5295 GTGTGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACT 5344

1502 ............................................... 1502

5345 ACACCCCAATGACATTTGAAGTGGTTAAAACCGGGACCTCTGAAGATGCC 5394
                                  |||:::   ::::::||||||
1503 ........................LysAsnProAlaLysGluAsp... 1509

5395 GTCGTGGAGTACTTGAAGTATCTGGCTATAGGCATTGAGAGGACATACAG 5444

1509 ............................................... 1509

5445 GGCGTTGCTTATGGCTAGAAATATTGCCGTCACTACCGCCGAAGGTGTTC 5494

1509 ............................................... 1509

5495 TGAAAGTACCTAATCAAGTTTATGAATCACTACCGGGCTTTCACGTTTAC 5544
                                                   :::
1510 ..........................................Phe 1510

5545 AAGTCGGGCACAGATCTCATTTTTCATTCAACACAAGACGGCTTGCGTGT 5594
     |||:::      |||      |||:::      :::   :::   :::||
1511 LysAlaIlePheAspHisAsnPheAsnIleMetGluGlyAsnAlaLysVa 1527

5595 GAGAGACCTACCGTACGTACTCATAGCTGAAAAAGGTATCTTTACCAAGG 5644
     |::::::::                        :::|||  :::|||:::::
1527 lLysGluPheAlaValSerSerAspGluTrpArgGly..MetTyrSerAs 1543

5645 GCAAAGATGTCGACGCGGTGGTAGCTTTGGGCGACAATCTGTTCGTATGC 5694
     ::::|||||||||  |||:::           ::::::||||||
1543 nGluArgCysArg.AlaIle...........GluGlnLeuPheAsnAsp 1555

5695 GACGATATACTGGTTTTCCACGATGCCATTAATTTGATAGGTGCACTGAA 5744
     ::::::   :::|||   :::      :::|||
1556 AsnAsnGluMetValArgArgProAspValAsn................. 1566

5745 AGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTAAGTCCTTCGAATATA 5794
                |||:::                              |
1567 ............GlyLeu............................L 1569

5795 AGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACGACGACGTTAGTGGAC 5844
     ||   ||||||   |||||||||:::|||||||||||||||:::::
1569 ysPheTyrAsnLysProProGlyAlaGlyLysThrThrThrIleAlaLys 1585

5845 GAATTCGTTAAGTCACCCAATAGCACAGCCACCATTACGGCT........ 5886
        :::   |||:::   ::::::   :::         |||
1586 LeuMetSerLysAspLeuLysAsnLysValLysCysLeuAlaLeuSerTy 1602

5887 ....AATGTGGGAAGTTCTGAGGACATAAATATG............GCGG 5920
         :::||||||      |||   |||:::              ::::
1602 rThrLysValGlyArgLeuGluLeuIleAspLysLeuLysLysAspGlyI 1619

5921 TGAAGAAGAGAGATCCGAATTTGGAAGGTCTCAACAGTGCTACCACAGTT 5970
     :::::|||   :::      ::::::      :::|||
1619 leGluLysProGluLysTyrValLysThrTyrAspSerPheLeuMetAsn 1635
```

```
 5971 AACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTT 6020
      |||:::::::::::::::::::::                  ::::::
 1636 AsnAspAsnIleLeuGluIleVal.................AsnLeuTy 1646

 6021 GGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAACTAGGCGTCT 6070
         |||||||||   |||||||||   |||   :::          :::
 1646 rCysAspGluValPheMetMetHisAlaGlyHisPheLeuThrLeuLeuT 1663

 6071 TCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATAAATCAGATA 6120
              :::   ::::::|||      :::||||||:::||||||||||
 1663 hrLysIleAlaTyrGlnAsnGlyTyrCysTyrGlyAspValAsnGlnIle 1679

 6121 CCATTCATAAACAGGGAGAAGGTGTTTAGGATGGATTGTGCTGTTTTTGT 6170
      ||||||||||||||:::                          :::::
 1680 ProPheIleAsnArgAspPro.......................TyrTh 1688

 6171 TCCA.......................AAGAAGGAAAGCGTTGTATACA 6196
      :|||                      :::|||:::::::::   |||
 1688 rProAlaTyrLeuSerArgGluPhePheArgLysGlnAspLeuAsnTyrA 1705

 6197 CTTCTAAATCGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCA 6246
        :::   :::||||||||||||||||||:::||||||||||||||||:::
 1705 spThrTyrThrTyrArgCysProLeuAspThrCysTyrLeuLeuSerAsn 1721

 6247 ATGACC..............................GTAAGGGGAAC 6263
      :::                                |||:::::::::
 1722 LeuLysAspGluMetGlyAsnIleIleTyrAlaGlyGlyValLysAsnVa 1738

 6264 GGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACCAGTAG 6313
      :::::::   |||                              |||::::
 1738 lAsnGluValTyr............................ProThrI 1745

 6314 TAAGATCGCTGTCCAAAAGGCCAATTGGAACCACTGATGACGTA...GCT 6360
      ::|||||||||:::         |||:::::::::   :::|||   :::
 1745 leArgSerLeuAsnLeuPheGlyIleAsnValValGlyGluValProVal 1761

 6361 GAAATAAACGCTGACGTGTACTTGTGCATGACCCAGTTGGAGAAGTCGGA 6410
      |||   ||||||      ||||||   :::||||||   ||||||   ::
 1762 GluTyrAsnAlaLys...TyrLeuThrPheThrGlnAspGluLysLeuAs 1777

 6411 TATGAAGAGGTCGTTGAAGGGAAAAGGAAAA...GAAACACCAGTGATGA 6457
      ::::::::|||   :::   ::::::|||      ::::::   |||   |
 1777 nLeuGlnArgHisIleAspSerGlnGlyGlyCysArgAsnAlaValSerT 1794

 6458 CAGTGCATGAAGCACAGGGAAAAACATTCAGTGATGTGGTATTGTTTAGG 6507
      |||||:::|||||||||||||   ||||||||||:::|||   |||   |||
 1794 hrValAsnGluAlaGlnGlyCysThrPheSerGluValAsnLeuValArg 1810

 6508 ACGAAGAAAGCCGATGACTCCCTATTCACTAAACAACCGCATATACTTGT 6557
            :::   |||:::   :::::::::         ::::::::::||
 1811 LeuValGlnPheAspAsnProValMetSerAspIleAsnGlnPheValVa 1827

 6558 TGGTTTGTCGAGACACACACGCTCACTGGTTTATGCCGCTCTGAGCTCAA 6607
      |::::::|||||||||||||   ::::::   |||   :::      |||:
 1827 lAlaIleSerArgHisThrThrThrPheLysTyrPheThrProHisSerA 1844

 6608 AGTTGGACGATAAGGTCGGCACATATATTAGCGACGCGTCACCTCAATCA 6657
      ::|||:::|||:::|||::::::   ||||||      |||   ||||||
 1844 rgLeuAsnAspArgValSerAsnAlaIleSer......SerLeuGlnSer 1858

 6658 GTATCCGACGCTTTGCTTCACACGTTC 6684
      |||||||||   :::|||      :::
 1859 ValSerAspPheValLeuLysAspTyr 1867

seq_name: sp_virus:Q9JGK3

seq_documentation_block:
ID   Q9JGK3      PRELIMINARY;      PRT;  1334 AA.
AC   Q9JGK3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   152-KDA PROTEIN.
OS   soil-borne wheat mosaic virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
OX   NCBI_TaxID=28375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAPANESE;
RA   Shirako Y., Suzuki N., French R.C.;
RT   "Similarity and divergence among viruses in the genus Furovirus.";
RL   Virology 270:201-207(2000).
DR   EMBL; AB033689; BAA94796.1; -.
SQ   SEQUENCE   1334 AA;  151950 MW;  288761E57551D0A7 CRC64;


alignment_scores:
           Quality:  486.00             Length:   1919
             Ratio:  0.606                Gaps:     82
 Percent Similarity:  41.793   Percent Identity: 19.854

alignment_block:
 US-09-301-906-3 x Q9JGK3   ..

 Align seg 1/1  to: Q9JGK3  from: 1  to: 1334

 1192 GCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGT 1241
      :::::::::   :::   :::|||:::|||||||||      ::::::::
    6 SerAsnIleLeuGlyIleIleAsnGluGluGluValValArgAlaAlaIl 22

 1242 TTCGAGTATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTT 1291
      :::::::   :::::|||                              |
   22 eAlaThrSerAlaThrArg.....................ThrAsnS 31

 1292 CCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGG......GAC 1335
      ||:::|||         :::|||:::::::::   ::::::      |||
   31 erGluLeuHisArgThrValCysGluGlnIleArgGluGlnPheLeuAsp 47

 1336 GCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTG......TTTCT 1379
      ::::::   :::::::|||   :::   :::   ||||||        ||
   48 ThrValGluAsnGlnLysGlnLysLysLysIleAspValArgArgAspLe 64

 1380 GAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGA 1429
      |      ::::::   :::   |||:::||||||:::||||||      |
   64 uThrGlnGluGlnLeuGlnThrLeuAsnGluLeuTyrProGluArgHisI 81

 1430 TACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATG 1479
      ||      |||          |||:::::::|||   ||||||   |||
   81 leValThrSerGlyCysGluArgGlyThrHisSerPheAlaAlaAlaSer 97

 1480 CGGAGCTGTTTCAATGGA...ATCTTTTCCAGGAGGTGTGGTAATGTGTG 1526
      |||:::      :::      ::::::||||||         |||:::
   98 ArgLysIleGluThrAspLeuLeuLeuSerArgMetProLysAsnAlaAs 114

 1527 CTTCTTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCATG 1576
          :::||||||||||||:::::::::   |||:::|||         :
  114 nValTyrAspIleGlyGlyAsnTrpAlaThrHisLeuLysArgLysAspT 131

 1577 TG...AACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAG 1623
      ::   :::   |||   |||   ||||||||||||   .::|||::::::
  131 hrArgLysValHisCysCysCysProValLeuAspPheArgAspAlaGln 147

 1624 CGGAGAATCAATGAGATCCTC.............TTTCTTTCCACAGCTGG 1661
      |||:::   ::::::   |||            |||||||||
  148 ArgLysThrThrArgTrpLeuSerValGluLysPheLeuSerGluLysAs 164

 1662 GGGA.............................................G 1666
        :::                                           :
  164 pSerIleSerGluHisCysGlyGlnLysValLeuGluIleGlnGluAspG 181

 1667 ATTCGTACGTGTCCAGTGACCTTCTAACTGAAGCG...GCTTCAAAGTCT 1713
      :::::   :::|||::::::|||          |||   |||   ::::::
  181 luAspArgIleSerAsnAsnLeuArgLysGlyAlaValAlaProGluAsp 197

 1714 GTGTCT.........TACTGTAGTCGAGAATCGCAGAACTGCGATTCTAG 1754
      ::::::         ||||||:::::::::   ::::::|||      ||
```

```
 198 LeuThrGlyArgTrpTyrCysGluAsnArgPheGluAspCysValTyrAr 214

1755 AGCCGATGCGGGTTTT.........ATGGTGGATGTGTACGATATATCCC 1795
     ||||:::   ::::::         :::   ::::::||||||||||||
 214 gAlaGluLysAlaTyrGlyMetAlaIleHisSerIleTyrAspIleSerL 231

1796 CGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGTGCGCTGGTTTTCGAC 1845
       ::::::::::::::::||||||:::::::|||            :::
 231 euAspAspLeuValAsnAlaMetGluGluLysArgIleLysTyrLeuVal 247

1846 ATAGCTCTTATGTTCCCCGTGGAGTTGTTGTACGGTAACGGTGAAGTTTA 1895
       ::::::::::|||   |||||||||:::   |||:::   :::
 248 GlyThrPheLeuPheSerValGluLeuPheLeuGlyLysLysArgGlyGl 264

1896 CTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAGGT.........GATT 1936
      |||   ::::::|||   :::   :::   :::|||             |
 264 uLeuThrSerValAspGlyPhePheGluIleAspGlyGlyAsnValLysT 281

1937 ACCTGGCCTACAATGTTGGTCAGTGTGGTGAGATGTATGAACATTCCTTC 1986
     ||       |||:::       :::||||||       |||:::|||::::::
 281 yrGlyPheTyrAspAspThrAsnCysGly......TyrArgHisAspLeu 295

1987 TCTAACGTAAGCGGGTTTTTCACCTTTTCTTATGTACGCACTTCGTCCGG 2036
     :::::::::       ::::::|||   ::::::|||:::::::::::||
 296 GlnGlnLeuMetGluTyrLeuThrLysThrPheValLysAlaLysGlyGl 312

2037 GAACGTGTTTAAGCTAGAGTATGAGGGATACCGTTGTGGTTACCATCATC 2086
     |:::|||||||   ||||||             |||              :
 312 ySerValPheTyrLeuGluLeuThrGluGlnArgGlyAspValMetPheP 329

2087 TCACTATGTGTAGGGCTCAGAAGTCACCTGGAACTGAGGTTACGTATAGG 2136
     ::::::::       |||                  |||||||:::       :::
 329 heSerLeuThrAspAla..............ThrGluAlaArgMetHis 340

2137 TCGTTGGTCCCGTCGTTCGTGGGCAAATCGCTGGTGTTCATACCTGTTGT 2186
     ::::::|||   :::          :::|||:::       ||||||:::
 341 GlyValValGluAsp.........GluSerPheLysCysIleProIleGl 354

2187 AGCTGGTTCTAGTGTG................................. 2202
      :::   :::|||||||
 354 uSerLysAspSerValValPheProLeuPheGluLeuAsnLysLysSerA 371

2203 ........TCCTTTAAGACAATAGTCCTCGATTCGGACTTTGTCGACAGG 2244
             :::|||       ::::::|||   ::::::||||||   |||
 371 spGluLeuGluPheThrGluValLeuLeuProLysSerPheValArgArg 387

2245 ATCTATTCCTACGCG.....................CTCAAC......AC 2267
          :::|||:::                      ||||||      ||
 388 ThrIleGluTyrThrAlaArgLeuLysThrAsnGlnLeuAsnProGluTh 404

2268 TATAGGGACATTCGAGAATAGAACGTTTGAGTATGCCGTTGGGGCGGTCA 2317
     |::::::::::::   :::   |||
 404 rValAsnSerTyrLeuThrSerThr..................... 412

2318 GGTCGCAAAAGACCCATGTCATTACAGGGAGT...CGCGTTGTCCACAGC 2364
          ::::::|||   :::|||   ||||||   :::   :::   :::
 413 .....AsnAsnThrValIleIleGlyGlySerAlaLysLysThrValGlu 427

2365 AAGGTTGATATTTCTCCTGATGATATGTGGGGTTTAGTTGTCGCTGTTAT 2414
     |||||||||                :::          ::::::   |||
 428 LysValAspAlaMetLeuIleProGlnIleThrThrThrLeuIleValTr 444

2415 GGCTCAGGCGATTAAGGATAGGGCGAAGAGTATTCGCTCCTATAACTTTA 2464
      ::::::   ::::::   |||   |||::::::
 444 pThrGluLeuMetAsnAlaArgGlnLysLysVal................ 455

2465 TAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCTCTTCTTTCAG 2514
                                  :::   :::|||   :::|||
 456 ..........................LeuAspArgLeuArgIleGln 462

2515 ACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATGCTAAGGCAAT 2564
           |||                      |||         |||::
 463 MetLysAspAsp.....................ValGlyPheMetAlaLe 472

2565 GGTGCACGATAACTTCAACGT....TTTGGAGACGCTTATGTCTATGCCC 2610
     ::::|||   :::|||:::      ||||||   :::      |||   :
 472 uAlaHisAlaSerPheSerLysMetPheGlyLysValSerPheTyrGlnL 489

2611 AGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCATTTGCAC 2660
     ::::::::::       ::::::|||:::      |||:::
 489 ysAlaLeuArgSerTyrAlaAsnTrpIleSerTyrCysHis..... 502

2661 TTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATATTTCGA 2710

 502 ................................................. 502

2711 AGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTCTTCTCT 2760
      ||| |||   :::   ::::::||||||      |||   |||
 503 .Gly.ThrAspAlaIleGluPheLysAsnValProLeuTyrVal..... 516

2761 AGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAAGACAGA 2810
          |||:::|||   :::|||:::   |||             ::
 517 ......GluValLysAspArgValLysMetTrpLysGlnHisAlaProAs 531

2811 AGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGTTCATAA 2860
     :   |||                                 ::::::|||   :
 531 nHisGly.............................PheSerPheAspL 538

2861 TGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTTGGCGGT 2910
     ::   ::::::
 538 euGluAspLeu..................................... 541

2911 TCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGCTCGTGG 2960

 541 ................................................. 541

2961 GGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCACAGGTT 3010

 541 ................................................. 541

3011 CACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGATGCTCTG 3060
                                    ||||||:::::::::
 542 ............................AspGluLysIleGlnLeuTyr 548

3061 GTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGAGACGTC 3110
        :::|||||||:::                              ::
 549 LysGluThrGluArg.................................Gl 554

3111 GCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTTGGAGTG 3160
     :|||   |||:::|||:::   |||:::::::   |||
 554 uArgLysAlaIleSerGluPheValValSerAspLys............. 566

3161 AAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAAGGTGCTT 3210
             |||:::|||:::::::::   |||   :::::::::
 567 .......LeuGlyAspLeuSerAlaAlaValCysGluLysGlu...... 578

3211 ATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATGGAAGTA 3260
                                    |||      ::::::   :::||
 579 ..........................AlaArgGluValSerValGluTy 586

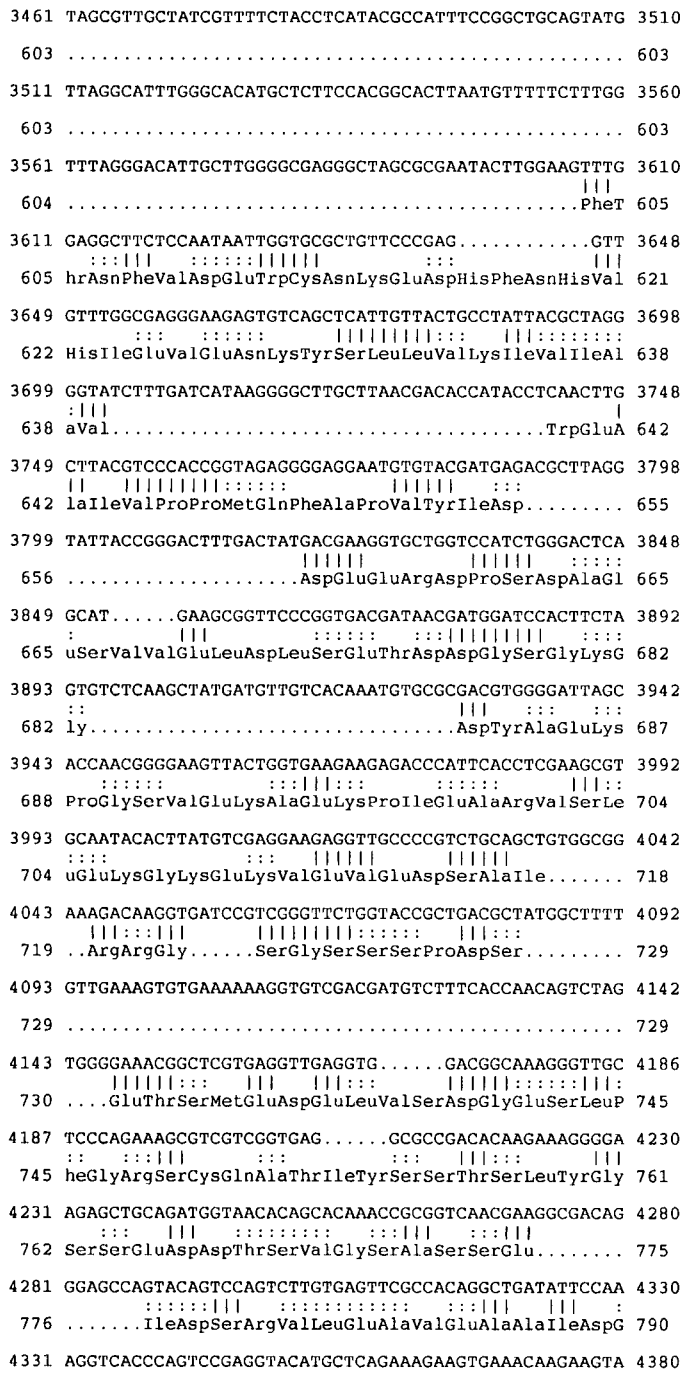
3261 TAGGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTAGTTACG 3310
     ||||                  ||||||
 586 rArg..................AspArgArg..................... 590

3311 AAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGCGGTTTA 3360
                                                  |||   :::
 591 ........................................ArgLysVal 593

3361 ACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGGGTTTTC 3410
        ::::::   :::   :::|||::::::
 594 GluAlaThrAspLeuArgAsnGlyGluVal..................... 603

3411 GTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAGTCCCCT 3460

 603 ................................................. 603
```


3461 TAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTGCAGTATG 3510
603 603
3511 TTAGGCATTTGGGCACATGCTCTTCCACGGCACTTAATGTTTTTCTTTGG 3560
603 603
3561 TTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGAAGTTTG 3610
604 PheT 605
3611 GAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAG.....GTT 3648
605 hrAsnPheValAspGluTrpCysAsnLysGluAspHisPheAsnHisVal 621
3649 GTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCTAGG 3698
622 HisIleGluValGluAsnLysTyrSerLeuLeuValLysIleValIleAl 638
3699 GGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCTCAACTTG 3748
638 aVal.....TrpGluA 642
3749 CTTACGTCCCACCGGTAGAGGGGAGGAATGTGTACGATGAGACGCTTAGG 3798
642 laIleValProProMetGlnPheAlaProValTyrIleAsp..... 655
3799 TATTACCGGGACTTTGACTATGACGAAGGTGCTGGTCCATCTGGGACTCA 3848
656 AspGluGluArgAspProSerAspAlaGl 665
3849 GCAT......GAAGCGGTTCCCGGTGACGATAACGATGGATCCACTTCTA 3892
665 uSerValValGluLeuAspLeuSerGluThrAspAspGlySerGlyLysG 682
3893 GTGTCTCAAGCTATGATGTTGTCACAAATGTGCGCGACGTGGGGATTAGC 3942
682 ly.....AspTyrAlaGluLys 687
3943 ACCAACGGGGAAGTTACTGGTGAAGAAGAGACCCATTCACCTCGAAGCGT 3992
688 ProGlySerValGluLysAlaGluLysProIleGluAlaArgValSerLe 704
3993 GCAATACACTTATGTCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGG 4042
704 uGluLysGlyLysGluLysValGluValGluAspSerAlaIle..... 718
4043 AAAGACAAGGTGATCCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTT 4092
719 ..ArgArgGly......SerGlySerSerSerProAspSer..... 729
4093 GTTGAAAGTGTGAAAAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAG 4142
729 729
4143 TGGGGAAACGGCTCGTGAGGTTGAGGTG......GACGGCAAAGGGTTGC 4186
730 GluThrSerMetGluAspGluLeuValSerAspGlyGluSerLeuP 745
4187 TCCCAGAAAGCGTCGTCGGTGAG......GCGCCGACACAAGAAAGGGGA 4230
745 heGlyArgSerCysGlnAlaThrIleTyrSerSerThrSerLeuTyrGly 761
4231 AGAGCTGCAGATGGTAACACAGCACAAACCGCGGTCAACGAAGGCGACAG 4280
762 SerSerGluAspAspThrSerValGlySerAlaSerSerGlu..... 775
4281 GGAGCCAGTACAGTCCAGTCTTGTGAGTTCGCCACAGGCTGATATTCCAA 4330
776 IleAspSerArgValLeuGluAlaValGluAlaAlaIleAspG 790
4331 AGGTCACCCAGTCCGAGGTACATGCTCAGAAAGAAGTGAAACAAGAAGTA 4380
790 luMet.....GluLysGluValSerAlaGluVal 799
4381 CCATTGGCGACTGTTTCGGGCGCCACGCCAATCGTCGATGAGAAACCCGC 4430
800 LeuAlaAlaValVal.....ValValThrGluSerSerGl 811
4431 CCCAAGTGTTACGACTCGTGGTGTGAAGATAATTGACAAGGGCAAGGCCG 4480
811 uProGluValSerGluValValVal.....GluAsnA 822
4481 TCGCTCATGTGGCTGAGAAAAAACAGGTACAAGTCGAGCAGCCCAAACAG 4530
822 laAlaThrValGluCysLysGluGlnValValValAspThrThrAspSer 838
4531 AGGAGTTTGACGATCAATGAAGGCAAGGCCGGTAAACAGCTTTGCATGTT 4580
839 GluSer.....AspGluGluArgAlaLys..... 846
4581 TAGAACGTGTTCCTGCGGTGTGCAGCTGGATGTGTACAACGAAGCGACTA 4630
847 SerTrpGly.....SerMetAlaGluGluGluSerA 857
4631 TCGCCACCAGGTTCTCAAACGCATTTACCTTTGTCGATAACTTGAAAGGG 4680
857 spAspThrPheTyrMetAsnThrMet...LeuIleSerAsnLysValGln 872
4681 AGGAGTGCGGTCTTTTTCTCAAAGCTGGGTGAGGGGTATACCTATAATGG 4730
873 ArgSerSerLeu..... 876
4731 TGGTAGCCATGTTTCATCAGGGTGGCCTCGTGCCCTAGAGGATATCTTAA 4780
877 ProLys..... 878
4781 CGGCAATTAAGTACCCAAGCGTCTTCGACCACTGTTTAGTGCAGAAGTAC 4830
878 878
4831 AAGATGGGTGGAGGCGTACCATTCCACGCTGATGACGAGGAGTGCTATCC 4880
879 GlnPr 880
4881 ATCAGATAACCCTATCTTGACGGTCAATCTCGTGGGGAAGGCAAACTTC. 4929
880 oAspPheAsnLysTyrAsnThrValGlnGlnLysAlaLysGlnGluTyrL 897
4930 TCGACTAAGTGCAGGAAGGGTGGTAAGGTCATGGTCATAAACGIA 4974
897 euTrpTyrLeuArgCysLys.....MetIleSerAsp... 907
4975 GCTTCGGGTGACTATTTTCTTATGCCTTGCGGTTTTCAAAGGACGCACTT 5024
908 ArgThrThrLe 911
5025 GCATTCAGTAAACTCCATCGACGAAGGGCGCATCAGTTTGACGTTCAGGG 5074
911 uArgSerIle..... 914
5075 CAACTCGGCGCGTCTTTGGTGTAGGCAGGATGTTGCAGTTAGCCGGCGGC 5124
914 914
5125 GTGTCGGATGAGAAGTCACCAGGTGTTCCAAACCAGCAACCACAGAGCCA 5174
915 IleGluAspHisLeuAsnGlyLeuTyrHisAsn..... 925
5175 AGGTGCTACCAGAACAATCACACCAAAATCGGGGGGCAAGGCTCTATCTG 5224
926 GlyAsnCysAspLeuProL 932
5225 AGGGAAGTGGTAGGGAAGTCAAGGGGAGGTCGACATACTCGATATGGTGC 5274

```
 932 ysAsnSer.....Cys 935

5275 GAACAAGATTACGTTAGGAAGTGTGAGTGGCTCAGGGCTGATAATCCAGT 5324
       ||||||:::::: ||||||:::|||
 936 PheLeuAspTyrThrLys.SerValGlyGly..... 945

5325 GATGGCTCTTGAACCTGACTACACCCCAATGACATTTGAAGTGGTTAAAA 5374
                                          :::||||||
 946 .....GluTrpLeuTyr 949

5375 CCGGGACCTCTGAAGATGCCGTCGTGGAGTACTTGAAGTATCTGGCTATA 5424
       |||||||||
 950 AlaLysProLeuLys..... 954

5425 GGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATATTGCCGT 5474
                  :::|||:::||| ||||||
 955 ..............AlaGlyHisCys.TyrGly..... 960

5475 CACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAATCAC 5524

 960 ..... 960

5525 TACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCATTCA 5574
                        ||||||          ||||||:::|
 961 ....................ValGly...........PheSerLeuA 966

5575 ACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAGCTGA 5624
     ||:::::: |||                          |||:::::
 966 snSerLys.Gly.........................LysIleGlyLy 973

5625 AAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCG..... 5661
     :      ::::::            ||||||||||||
 973 sCysGluLeuLeuLysLeuMetTrpAspValAspAlaArgGlyGlnProL 990

5662 ..........................GTGGTAGCTTTGGGCGAC 5679
                                        |||   |||:::|||
 990 euGluArgProPheAsnThrLysMetPheGlnPheValLeuLeuAsnAsp 1006

5680 AATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGCCATTAATTT 5729
           |||:::   ::::::::::::::|||:::          |||||
1007 LeuThrPheLeuMetAsnGluMetIleIlePheArg.........AsnLe 1020

5730 GATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCGAATCGTTTA 5779
     |      :::|||:::                          |||      |
1020 uGlnAspThrLeuGlnArgLys.................GluArgThrL 1031

5780 AGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGCGGTAAGACG 5829
     ||:::   :::             ::::::   ||||||   |||||||:::
1031 ysGlnAlaSerIleValLeuLysAspGlyValProGlyCysGlyLysSer 1047

5830 ACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCACAGCCACCAT 5879
     |||   :::::::::            :::   |||            ::
1048 ThrTrpIleLeuAsn............AsnAlaAsnTyrIleLysAspVa 1060

5880 TACGGCTAATGTGGGA......AGTTCTGAGGACATAAATATGGCGGTGA 5923
     ::::   :::||||||      ::::::||||||::::::
1060 lValIleSerValGlyLysGluAlaLysGluAspLeuLysGluLysPheM 1077

5924 AGAAGAGAGATCCGAATTTGGAAGGT...CTCAACAGTGCTACCACAGTT 5970
       |||:::          |||:::   |||:::          ||||||
1077 etLysLysTyrLysCysThrGluSerGluLeuGlyArgIleArgThrVal 1093

5971 AACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAAAAGGGTTTT 6020
     :::|||            ||||||||||          |||::::::
1094 AspSer............TyrLeuMetHisAspCysGlyLysLysLeuAr 1106

6021 GGTGGATGAGGTGCAC..............ATGATGCATCAAGGCTTAC 6055
       :::      ||||||              |||   |||   |||   :
1106 gAlaThrThrValHisPheAspGluAlaLeuMetThrHisAlaGlyAlaV 1123

6056 TACAACTAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGA 6105
     ::   :::   :::         ||||||   :::   :::      |||
1123 alTyrPheCysAlaAspLeuLeuGlyAlaArgLysValIleCysGlnGly 1139

6106 GACATAAATCAGATACCATTCATAAACAGGGAGAAGGTGTTTAGGATGGA 6155
     |||   :::|||||||||||||:::||||||   :::   ::::::::::
1140 AspSerGlnGlnIleProPheValAsnArgValGluSerIleLysLeuGl 1156

6156 TTGTGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAAT 6205
     :   |||   :::|||       |||   :::::::::            :
1156 nPheAlaLysLeuValIleAspLysThrAspLeuIle......ArgMetT 1171

6206 CGTACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTA 6255
     ::||||||   |||:::||||||::::::  |||:::
1171 hrTyrArgSerProIleAspValAlaHisTyrLeuAsn..... 1183

6256 AGGGGAACGGAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAG..... 6300
                          |||   :::      ::::::|||:::
1184 ..................TyrLysSerPheTyrThrGlyGlyArgIleTh 1194

6301 ....GACAAACCAGTAGTAAGATCGCTGTCC.................A 6328
           :::   ||||||||||||:::|||                 .
1194 rThrLysAsnGluValValArgSerMetSerValValGlyProArgAsnV 1211

6329 AAAGGCCAATTGGAACCACTGATGACGTAGCTGAAATAAACGCTGACGTG 6378
       ||||||:::   ::::::   :::|||      :::   :::
1211 alArgProMetThrSerValTyrSerValProTyrValProGlyValGln 1227

6379 TACTTGTGCATGACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAA 6428
     ||||||   :::|||||||   ||||||:::|||:::   ::::::|||::
1228 TyrLeuThrPheThrGlnThrGluLysAspAspLeuPheLysAlaLeuAr 1244

6429 GGGAAAAGGAAAAGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAA 6478
     ::::||||||   ::::::   |||   |||||||||||:::|||||||
1244 gSerLysGly...HisValAsnValAsnThrValHisGluThrGlnGlyL 1260

6479 AAACATTCAGTGATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCC 6528
     ||||||||:::||||||:::|||   |||   |||   ::::::::::::
1260 ysThrPheAspAspValIleLeuValArgLeuLysThrThrGluAsnGlu 1276

6529 CTATTC............ACTAAACAACCGCATATACTTGTTGGTTTGTC 6566
     ::::::            ::::::::::|||:::   :::||||||||||::
1277 IleTyrProGlyGlyArgAsnSerLysProTyrThrIleValGlyLeuTh 1293

6567 GAGACACACACGCTCACTGGTT...TATGCCGCTCTGAGCTCAAAGTTG 6612
     :||||||   ||||||||||||   |||:::|||:::::::::::::|||
1293 rArgHisArgArgSerLeuValTyrTyrThrAlaIleGluAspArgLeu 1309
```

seq_name: sp_virus:Q9WIE3

seq_documentation_block:

```
ID   Q9WIE3      PRELIMINARY;      PRT;  1350 AA.
AC   Q9WIE3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   REPLICASE.
OS   Chinese wheat mosaic virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
OX   NCBI_TaxID=83544;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YANTAI, CHINA;
RX   MEDLINE=99281891; PubMed=10355760;
RA   Diao A., Chen J., ye R., Zheng T., Yu S., Antoniw J.F., Adams M.J.;
RT   "Complete sequence and genome properties of Chinese wheat mosaic
RT   virus, a new furovirus from China.";
RL   J. Gen. Virol. 80:1141-1145(1999).
DR   EMBL; AJ012005; CAB41770.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   1350 AA;  153005 MW;  F8D84D51BD426B1F CRC64;
```

```
alignment_scores:
            Quality:  484.00              Length:   1958
              Ratio:  0.598                Gaps:     82
 Percent Similarity:  41.318   Percent Identity: 19.408

alignment_block:
 US-09-301-906-3 x Q9WIE3   ..

 Align seg 1/1  to: Q9WIE3  from: 1  to: 1350

    1192 GCCTCCGTCGCCAATGCCCTTAATCAAGAGGAGGTCTATTCATCGGTCGT 1241
         :::|||:::   :::   :::::::::|||:::|||      ::::::::
       6 SerSerIleLeuGlyIleIleSerGluGluAspValIleArgAlaAlaIl 22

    1242 TTCGAGTATCACCAATAGACTGGTATTAAGGGACCAATCGGCATTGCTTT 1291
         :|||:::   ::::::::::::                           |
      22 eSerThrSerAlaThrLysPheGly.......................S 31

    1292 CCCATTTGGACACGAAATTGTGCGATATGTTTTCTCAAAGG......GAC 1335
         ||:::|||   :::   :::||||||      ::::::         |||
      31 erGlnLeuHisSerThrValCysAspHisValLysGluThrTyrLeuAsp 47

    1336 GCAATGATTCGCGAAAAACCCTCACATAGGTGCGATGTG......TTTCT 1379
         ||||||   |||:::|||   :::   :::   ||||||         ||
      48 AlaMetGluArgGlnLysThrLysLysLysIleAspValArgArgAspLe 64

    1380 GAAGCCGCGGGAAAGGGAGAAGCTGAGGGAACTCTTTCCAGAGCTTTCGA 1429
         |:::   ::::::   :::   |||::::::|||:::||||||      |
      64 uSerGlnGluGlnLeuGlnLeuLeuAsnAspLeuTyrProGluArgHisI 81

    1430 TACAGTTCTCCGACTCGGTCAGGAGTAGTCACCCATTCGCTAATGCCATG 1479
         ||      |||:::      |||::::::|||   ||||||   |||
      81 leValSerSerAsnCysGluArgGlyThrHisSerPheAlaAlaAlaSer 97

    1480 CGGAGCTGTTTCAATGGAATCTTTTCCAGGAGGTGTGGTAATGTGTGCTT 1529
         |||:::      :::   ::::::      |||      :::      ::
      98 ArgLysIleGluThrAspLeuLeuLeuSerArgIleProLysLysSerTr 114

    1530 C...TTCGATATTGGGGGGAGCTTCACGTATCATGTCAAAGCTGGCCAT. 1575
         :   :::||||||||||||:::::::::   |||:::|||   :::
      114 pValTyrAspIleGlyGlyAsnTrpAlaThrHisIleLysArgAsnAspG 131

    1576 ..GTGAACTGTCATGTATGCAATCCAGTCCTAGACGTTAAAGATGTGAAG 1623
           :::   |||   |||   |||:::::|||:::::|||
      131 lyArgArgValHisCysCysCysProThrIleAspIleArgAspSerAla 147

    1624 CGGAGAATCAATGAGATCCTCTTTCTTTCCACAGCTGGGGGAGATTCGTA 1673
         |||:::                  :::|||   |||:::   ::::::||
      148 ArgLysThr..............ValSerTrpAlaSerIleGluLysTy 159

    1674 CGTG......TCCAGTGACCTTCTAACTGAAGCGGCTTCA......AAGT 1711
         |:::      ::::::::::::   ||||||:::::::::      |||
      159 rLeuAspGluLysAspGluIleProThrGluValGlyGluArgIleLysA 176

    1712 CTGTGTCT.......................................... 1719
           |||:::
      176 rgValGlnAlaAspGluIleLysIleTyrGluAsnLeuLysAsnLysArg 192

    1720 ............................TACTGTAGTCGAGAATCGCA 1739
                                     ||||||::::::::::   ::
      193 AlaGlnProGluAspLeuAspGlyLysTrpTyrCysGlyAsnLysPheGl 209

    1740 GAACTGCGATTCTAGAGCCGATGCGGGTTTTATGGTG.........GATG 1780
         ::::|||      ||||||:::   ::::::   :::         ::::
      209 uGluCysValPheArgAlaGluArgAlaTyrAlaMetAlaIleHisSerI 226

    1781 TGTACGATATATCCCCGCAGCAGGTAGCAGAGGCTATGGATAAGAAGGGT 1830
         ::||||||||||:::   ::::::::::|||:::|||::::::::|||
      226 leTyrAspIleAspLeuMetAspLeuAlaAsnAlaLeuGluGluLysArg 242

    1831 GCGCTGGTTTTCGACATAGCTCTTATGTTCCCCGTGGAGTTGTTGTACGG 1880
               :::::::::   ::::::::::|||   |||:::||||||   ||
      243 IleLysIleMetSerGlyThrPheLeuPheSerValAspLeuLeuLeuGl 259

    1881 TAACGGTGAAGTTTACTTGGAAGAACTCGATACGTTGGTGAAGAGGGAAG 1930
         |:::   :::      |||      :::|||   :::   :::   ||||
      259 yLysLysSerGlySerLeuProThrMetAspGlyPhePheGluValGluG 276

    1931 GTGATTACCTGGCCTAC.................AATGTTGGTCAGTGT 1962
         ||||||||:::   |||                  |||   |||
      276 lyAspTyrValLysTyrGlyPheLysAsnAspThrAsnProGly...... 290

    1963 GGTGAGATGTATGAACATTCCTTCTCTAACGTAAGCGGGTTTTTCACCTT 2012
                |||:::|||:::::::|||::::::      :::::::|||
      291 .........TyrLysHisLysLeuSerGlnLeuMetLysTyrLeuThrLy 304

    2013 TTCTTATGTACGCACTTCGTCCGGGAACGTGTTTAAGCTAGAGTATGAGG 2062
         ::::::|||   :::::::::|||:::::::::   ||||||
      304 sThrPheValValAlaLysGlyGlyThrIleTyrTyrLeuGluLeuThrG 321

    2063 GATACCGTTGTGGTTACCATCATCTCACTATGTGTAGGGCTCAGAAGTCA 2112
         |||                  :::||||||      |||
      321 luGlnArgGlyAspValMetPhePheThrMetThrAspAla......... 334

    2113 CCTGGAACTGAGGTTACGTATAGGTCGTTGGTC......CCGTCGTTCGT 2156
               ||||||:::      :::::::::|||         ||||||
      335 ......ThrGluAlaArgMetAsnGlyValValAlaAspGluSerPheLy 349

    2157 G..............GGCAAATCGCTGGTGTTCATACCTGTTGTA.... 2187
                        :::||||||   |||   :::|||:::
      349 sCysIleProIleAspAsnLysSerGluValValPheProIlePheGluV 366

    2188 ........GCTGGTTCTAGTGTGTCCTTTAAGACAATAGTCCTCGATTCG 2229
                    :::|||:::   |||:::   ::::::|||::::::
      366 alAspLysGlnThrAspSerLeuIlePheSerGluValLeuLeuSerLys 382

    2230 GACTTTGTCGACAGGATCTATTCCTACGCG..................... 2259
         |||||||||:::|||      :::::::::
      383 AspPheValGlnArgAlaIleGluPheThrGlyArgLeuLysProAsnGl 399

    2260 ................CTCAACACTATAGGGACATTCGAGAATAGAACGT 2293
                        :::||||||      |||      |||:::|||:
      399 nLeuThrProAspAsnValAsnThrTyrLeuThrSerThrAsnAsnThrI 416

    2294 TTGAGTATGCCGTTGGGGCGGTCAGGTCGCAAAAGACCCATGTC...... 2337
         ::      :::   |||:::   :::::::::   |||   |||
      416 leIleIleGly...GlySerSerLysLysAsnThrThrLysValAspAla 431

    2338 ...............ATTACAGGGAGTCGCGTTGTCCACAGCAAGGTTGA 2372
                         ||||||   :::   ||||||   :::
      432 ThrLeuIleGlnGlnIleThrThrThrLeuValValTrpThr........ 445

    2373 TATTTCTCCTGATGATATGTGGGGTTTA............GTTGTCGCTG 2410
                        ::::::   :::|||            |||:::
      446 ............GluLeuMetAsnLeuArgGlnLysArgValLeuGluL 458

    2411 TTATGGCTCAGGCGATTAAGGATAGGGCG......AAGAGTATTCGCTCC 2454
           :::         :::||||||   :::      ||||||:::
      458 ysLeuArgLeuGlnMetLysAspAspValAspPheLysSerLeuAlaHis 474

    2455 TATAACTTTATAAAAGCCAGTGAGGGGAGTCTCGCCGGGGTCTTCAAGCT 2504
               |||:::|||:::      |||:::::::::
      475 ThrAlaPheLeuLysValPhe...GlyLysValSer.............. 485

    2505 CTTCTTTCAGACCGTAGGCGATTGTTTTTCGAACGCAGTCTCCGTCTATG 2554
         ::::::|||   :::      ||||||:::|||   :::|||
      486 .TyrTyrGlnArgAlaLeuArgCysPheAlaAsnTrpLeuSer....... 499

    2555 CTAAGGCAATGGTGCACGATAACTTCAACGTTTTGGAGACGCTTATGTCT 2604
                |||||
      500 ........TyrValHis.............................. 502

    2605 ATGCCCAGAGCGTTCATCCGTAAAGTACCTGGGTCTGTTGTTGTTACCAT 2654

      502 .............................................. 502
```

```
2655 TTGCACTTCTGGAGCTTCAGACAGGTTGGAGCTCAGGGGTGCCTTTGATA 2704
                :::::::|||   ::::::::::::
 503 .............GlyThrAspAlaIleGlnPheHisAsp.......... 511

2705 TTTCGAAGGAGACCTTCGGTAGGAAACTGAAGAATAGTCGCTTGCGCGTC 2754
                                                      |||
 512 ..................................................Val 512

2755 TTCTCTAGGGCTATCGTGGAAGATTCAATTAAGGTCATGAAGGCAATGAA 2804
           |||   |||   |||   :::|||:::   |||:::
 513 ProLeuTyrAlaGluValThrAspArgValLysLeuTrpLysThrHisAl 529

2805 GACAGAAGATGGAAAACCCCTGCCAATTACTGAAGATTCTGTATATGCGT 2854
     :::::::|||      |||   :::   ||||||           :
 529 aProAsnGlnGlyPheValLeuAspMet...GluAspLeuAspValLysI 545

2855 TCATAATGGGGAACGTTTCTAACGTCCACTGTACGAGGGCAGGTCTTCTT 2904
     ::   |||   :::||||||
 545 leLysMetHisGluValSer.............................. 551

2905 GGCGGTTCGAAAGCGACCGTGGTTTCGAGTGTTTCTAAGGGTTTGGTAGC 2954

 551 ..................................................... 551

2955 TCGTGGGGCTGCGACGAAGGCCTTTTCTGGCATTACGTCGTTCTTTTCCA 3004

 551 ..................................................... 551

3005 CAGGTTCACTATTCTACGACCGCGGTTTAACTGAAGATGAAAGGCTTGAT 3054
                                  |||   |||:::   |||
 552 ..............................GluArgGluLysArgAsp 557

3055 GCTCTGGTGCGCACAGAGAATGCTATAAACTCACCGGTGGGCATACTGGA 3104

 557 ..................................................... 557

3105 GACGTCGCGCGTAGCTGTGAGCAAGGTCGTAGCTGGAACGAAAGAATTTT 3154
      :::||||||      ||||||     .:::   |||      ::::::
 558 .ValSerArgCysIleValSerGlyLysLeu...GlyGluLeuGlnIleH 573

3155 GGAGTGAAGTTTCCTTAAATGACTTCACCACTTTCGTATTGCGGAATAAG 3204
       ||||||   |||   |||:::                    |||  |
 573 isSerGluSerSerAsnAsnGluGluGlnIle............Asn..G 585

3205 GTGCTTATCGGGATATTCGTGGCGTCTTTGGGTGCGGCCCCAATTGCATG 3254
     ||   |||||||||
 585 lyAspTyrArgAsp..................................... 589

3255 GAAGTATAGGCGCGGAATTGCGGCTAACGCTAGAAGGTACGCGGGCAGTA 3304

 589 ..................................................... 589

3305 GTTACGAAACTCTAAGCTCGTTAAGTTCACAAGCCGCCGGTGGTTTACGC 3354
                                   |||||||||      |||
 590 ...............................SerArgArgArgThrThr. 595

3355 GGTTTAACCTCTAGCACAGTATCCGGTGGATCTTTAGTCGTGCGAAGAGG 3404

 595 ..................................................... 595

3405 GTTTTCGTCGGCGGTGACCGTCACTAGGGCGACCGTAGCTAAACGTCAAG 3454

 595 ..................................................... 595

3455 TCCCCTTAGCGTTGCTATCGTTTTCTACCTCATACGCCATTTCCGGCTGC 3504

 595 ..................................................... 595

3505 AGTATGTTAGGCATTTGGGCACATGCTCTTCCACGGCACTTAATGTTTTT 3554

 595 ..................................................... 595

3555 CTTTGGTTTAGGGACATTGCTTGGGGCGAGGGCTAGCGCGAATACTTGGA 3604
```

```
          |||:::|||:::      ||||||        ::::::|||::
 596 ......PheGluAspLeuLeuGluGlyGluValAlaThrAsnPheLeuAs 610

3605 AGTTTGGAGGCTTCTCCAATAATTGGTGCGCTGTTCCCGAGGT....... 3647
     :   |||                |||:::::::::
 610 pAspTrp.....................CysGluLysAsnAspHisP 619

3648 ..TGTTTGGCGAGGGAAGAGTGTCAGCTCATTGTTACTGCCTATTACGCT 3695
       :::::::|||:::::::::::|||            :::   ||||||
 619 heAsnPheSerArgAspAspAlaGln............SerLysTyrAla 631

3696 AGGGGTATCTTTGATCATAAGGGGCTTGCTTAACGACACCATACCTCAAC 3745
     |||:::    ::::::::::|||:::   :::       :::|||   :
 632 TrpGlyLeu.LysLeuLeuLysGlyIleTrpGluPheLeuLeuProProM 648

3746 TTGCTTACGTCCCACCGGTAGAGGGGAGGAATGTGTAC......GATGAG 3789
     ::   :::::::|||                |||||      ||||||
 648 etAspPheAlaPro.................ValTyrValAspAspGlu 658

3790 ACGCTTAGGTATTACCGGGACTTTGACTATGACGAAGGTGCTGGTCCATC 3839
          |||      |||            ::::||      :::   ||
 659 GlnAlaArg...LeuArgMetValArgValAsnGluArgArgAsnThrSe 674

3840 TGGGACTCAGCATGAAGCGGTTCCCGGTGACGATAACGATGGATCCACTT 3889
     |   |||                  |||   |||:::|||   ::::::
 674 rAspThr..................GlyAlaAspGlyAspValAlaAlaC 685

3890 CTAGTGTCTCAAGC.........................TATGATGTT 3912
       :::|||::::::                         ||||||
 685 ysGluValAlaAlaAspPheSerLysAlaMetAspThrLeuValAspVal 701

3913 GTCACAAATGTG......CGCGACGTGGGGATTAGCACCAACGGGGAAGT 3956
     |||   ::::::      :::   :::|||:::
 702 ValLysLysMetAspGlnLysLysLeuGlyLeu................. 712

3957 TACTGGTGAAGAAGAGACCCATTCACCTCGAAGCGTGCAATACACTTATG 4006
                                                       |
 713 .....................................................V 713

4007 TCGAGGAAGAGGTTGCCCCGTCTGCAGCTGTGGCGGAAAGACAAGGTGAT 4056
     |||||:::      |||   ::::::||||||:::|||   :::
 713 alGluLysValLysAlaSerAlaValAlaValValGluLeuSerLysThr 729

4057 CCGTCGGGTTCTGGTACCGCTGACGCTATGGCTTTTGTTGAAAGTGTGAA 4106
           :::|||::::::   |||:::   :::       :::   ::
 730 LysProValThrGlyAsnSerCysAlaValAspLeuTrpAlaAspPheGl 746

4107 AAAAGGTGTCGACGATGTCTTTCACCAACAGTCTAGTGGGGAAACGGCTC 4156
     :|||:::::|||||              :::::::::          :::
 746 uLysAsnLeuAspAsp............AspAspGluProGlyCysGlyI 759

4157 GTGAGGTTGAGGTGGACGGCAAAGGGTTGCTCCCAGAAAGCGTCGTCGGT 4206
       :::|||   :::   ||||||   :::       :::::|||
 759 leSerValValThrLysGlyLysGluIleCysGlnAspAsnVal...... 773

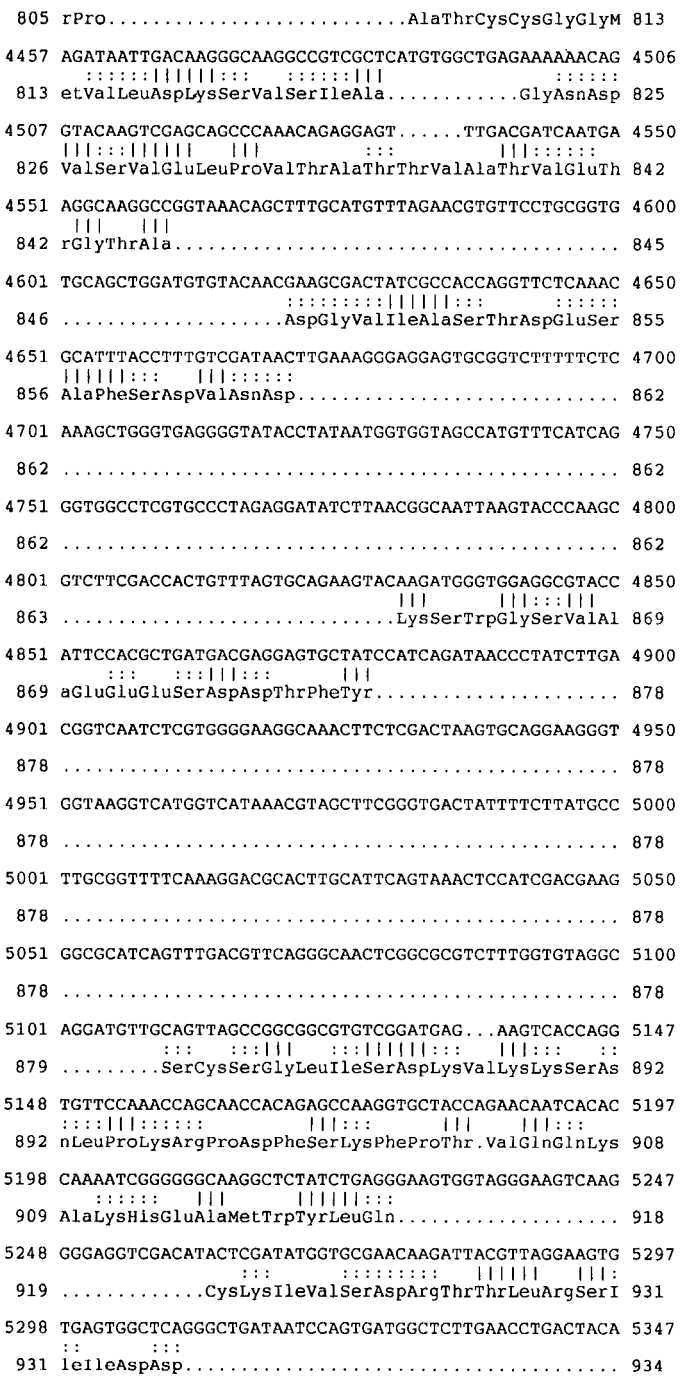
4207 GAGGCGCCGACACAAGAAAGGGGAAGAGCTGCAGATGGTAACACAGCACA 4256
        :::|||:::                          |||::::::::::
 774 ...GlyProVal..............LeuLeuCysGlySerSerSerCy 784

4257 AACCGCGGTCAACGAAGGCGACAGGGAGCCAGTACAGTCCAGTCTTGTGA 4306
     ::::::|||:::|||   ::::::|||
 784 sSerSerValSerGluValGluLysGlu..................... 793

4307 GTTCGCCACAGGCTGATATTCCAAAGGTCACCCAGTCCGAGGTACATGCT 4356
               :::|||:::   :::||||||:::
 794 ...........ThrAspValValSerValThrAsp............... 801

4357 CAGAAAGAAGTGAAACAAGAAGTACCATTGGCGACTGTTTCGGGCGCCAC 4406
                                                 ||||||   ::
 802 ..................................SerGlyAsnSe 805

4407 GCCAATCGTCGATGAGAAACCCGCCCCAAGTGTTACGACTCGTGGTGTGA 4456
     :|||                          :::|||      |||
```


805 rPro.....AlaThrCysCysGlyGlyM 813
4457 AGATAATTGACAAGGGCAAGGCCGTCGCTCATGTGGCTGAGAAAAAACAG 4506
813 etValLeuAspLysSerValSerIleAla.....GlyAsnAsp 825
4507 GTACAAGTCGAGCAGCCCAAACAGAGGAGT.....TTGACGATCAATGA 4550
826 ValSerValGluLeuProValThrAlaThrThrValAlaThrValGluTh 842
4551 AGGCAAGGCCGGTAAACAGCTTTGCATGTTTAGAACGTGTTCCTGCGGTG 4600
842 rGlyThrAla..... 845
4601 TGCAGCTGGATGTGTACAACGAAGCGACTATCGCCACCAGGTTCTCAAAC 4650
846 AspGlyValIleAlaSerThrAspGluSer 855
4651 GCATTTACCTTTGTCGATAACTTGAAAGGGAGGAGTGCGGTCTTTTTCTC 4700
856 AlaPheSerAspValAsnAsp..... 862
4701 AAAGCTGGGTGAGGGGTATACCTATAATGGTGGTAGCCATGTTTCATCAG 4750
862 862
4751 GGTGGCCTCGTGCCCTAGAGGATATCTTAACGGCAATTAAGTACCCAAGC 4800
862 862
4801 GTCTTCGACCACTGTTTAGTGCAGAAGTACAAGATGGGTGGAGGCGTACC 4850
863 LysSerTrpGlySerValAl 869
4851 ATTCCACGCTGATGACGAGGAGTGCTATCCATCAGATAACCCTATCTTGA 4900
869 aGluGluGluSerAspAspThrPheTyr..... 878
4901 CGGTCAATCTCGTGGGGAAGGCAAACTTCTCGACTAAGTGCAGGAAGGGT 4950
878 878
4951 GGTAAGGTCATGGTCATAAACGTAGCTTCGGGTGACTATTTTCTTATGCC 5000
878 878
5001 TTGCGGTTTTCAAAGGACGCACTTGCATTCAGTAAACTCCATCGACGAAG 5050
878 878
5051 GGCGCATCAGTTTGACGTTCAGGGCAACTCGGCGCGTCTTTGGTGTAGGC 5100
878 878
5101 AGGATGTTGCAGTTAGCCGGCGGCGTGTCGGATGAG...AAGTCACCAGG 5147
879 SerCysSerGlyLeuIleSerAspLysValLysLysSerAs 892
5148 TGTTCCAAACCAGCAACCACAGAGCCAAGGTGCTACCAGAACAATCACAC 5197
892 nLeuProLysArgProAspPheSerLysPheProThr.ValGlnGlnLys 908
5198 CAAAATCGGGGGGCAAGGCTCTATCTGAGGGAAGTGGTAGGGAAGTCAAG 5247
909 AlaLysHisGluAlaMetTrpTyrLeuGln..... 918
5248 GGGAGGTCGACATACTCGATATGGTGCGAACAAGATTACGTTAGGAAGTG 5297
919 CysLysIleValSerAspArgThrThrLeuArgSerI 931
5298 TGAGTGGCTCAGGGCTGATAATCCAGTGATGGCTCTTGAACCTGACTACA 5347
931 leIleAspAsp..... 934
5348 CCCCAATGACATTTGAAGTGGTTA...AAACCGGGACCTCTGAAGATGCC 5394
935 HisLeuArgGlyMetPheHisAsnGlyThrCysGluLeuPr 948
5395 GTCGTGGAGT.....ACTTGAAGTATCTGGC 5420
948 oLysAspSerAlaPheLeuAspTyrThrAlaAspAsnCysGlyThrTrpM 965
5421 TATAGGCATTGAGAGGACATACAGGGCGTTGCTTATGGCTAGAAATATTG 5470
965 etTyrGlyLysProSerArgPheGlyHis.SerTyrGlyValGlyPheSe 981
5471 CCGTCACTACCGCCGAAGGTGTTCTGAAAGTACCTAATCAAGTTTATGAA 5520
981 rLeuAspThrArgGlnArgValSerLysCysGlu..... 992
5521 TCACTACCGGGCTTTCACGTTTACAAGTCGGGCACAGATCTCATTTTTCA 5570
993 LeuValLysLeuMetTrpAsnHisAspSerArgGlyGln 1005
5571 TTCAACACAAGACGGCTTGCGTGTGAGAGACCTACCGTACGTACTCATAG 5620
1006 ValAsn.GlnLysProValAsnThrArgAlaPheGlnTyrLeuLeuLeuS 1022
5621 CTGAAAAAGGTATCTTTACCAAGGGCAAAGATGTCGACGCGGTGGTAGCT 5670
1022 erGluLeuSer..... 1025
5671 TTGGGCGACAATCTGTTCGTATGCGACGATATACTGGTTTTCCACGATGC 5720
1026 PheMetMetAsnGluMetIleIleTyrArg..... 1035
5721 CATTAATTTGATAGGTGCACTGAAAGTCGCTCGATGCGGCATGGTGGGCG 5770
1036 AsnLeuGlnGlnValMetArgLysArg.....G 1045
5771 AATCGTTTAAGTCCTTCGAATATAAGTGCTATAATGCTCCCCCAGGTGGC 5820
1045 luArgSerLysGlnAlaArgIleThrLeuArgAspGlyValProGlyCys 1061
5821 GGTAAGACGACGACGTTAGTGGACGAATTCGTTAAGTCACCCAATAGCAC 5870
1062 GlyLysSerThrTrpIleLeuAsn.....AsnAlaAsnProTh 1074
5871 AGCCACCATTACGGCTAATGTGGGA.....AGTTCTGAGGACATAAATA 5914
1074 rLysAspMetIleLeuSerValGlyLysGluAlaThrGluAspLeuLysG 1091
5915 TGGCGGTGAAGAAGAGAGATCCGAATTTGGAAGGT...CTCAACAGTGCT 5961
1091 luLysPheMetLysLysHisArgCysValGluSerAspLeuLysArgIle 1107
5962 ACCACAGTTAACTCCAGGGTGGTAAACTTTATCGTCAGGGGAATGTATAA 6011
1108 ArgThrValAspSerPheLeuMetHisAspTyrAspLysTyrArgAlaAl 1124
6012 AAGGGTTTTGGTGGATGAGGTGCACATGATGCATCAAGGCTTACTACAAC 6061
1124 aThrValHisPheAspGluAlaLeuMetAlaHisAlaGlyMetValTyrP 1141
6062 TAGGCGTCTTCGCAACCGGCGCGTCGGAAGGCCTCTTTTTTGGAGACATA 6111
1141 heCysAlaAspIleLeuGlyAlaLysLysValIleCysGlnGlyAspSer 1157
6112 AATCAGATACCATTCATAAACAGG...GAGAAGGTGTTTAGGATGGATTG 6158
1158 GlnGlnIleProPheIleAsnArgValGluSerIleThrLeuGlnTyrSe 1174
6159 TGCTGTTTTTGTTCCAAAGAAGGAAAGCGTTGTATACACTTCTAAATCGT 6208
1174 rLysLeuValIleAspGluThrGluHisValArgLeuThr.....T 1188

```
6209 ACAGGTGTCCGTTAGATGTTTGCTACTTGTTGTCCTCAATGACCGTAAGG 6258
     |||||   |||:::||||||:::::   |||::::::
1188 yrArgSerProValAspValAlaHisTyrLeuThrLys..... 1200

6259 GGAACGGAAAAAGTGTTACCCTGAAAAGGTCGTTAGCGGTAAGGACAAACC 6308
       ::::::   |||           |||:::   |||:::   ||
1201 ......LysSerTrpTyrSerGlyGlyArgValThrThrLysAsn...Pr 1214

6309 AGTAGTAAGATCGCTGTCCAAA.................AGGCCAATTG 6340
     ||||:::||||||::::::                    :::|||:::
1214 oValMetArgSerMetLysThrValGlyProArgAspValLysProMetT 1231

6341 GAACCACTGATGACGTAGCTGAAATAAACGCTGACGTGTACTTGTGCATG 6390
       ::::::      |||      ::::::         ||||||   :::
1231 hrSerValHisCysValProTyrPheLysAspAlaGlnTyrLeuThrPhe 1247

6391 ACCCAGTTGGAGAAGTCGGATATGAAGAGGTCGTTGAAGGGAAAAGGAAA 6440
     ||||||   ||||||:::|||:::   :::::::|||::::::||||||
1248 ThrGlnSerGluLysThrAspLeuTyrLysAlaLeuArgAsnLysGlyPr 1264

6441 AGAAACACCAGTGATGACAGTGCATGAAGCACAGGGAAAAACATTCAGTG 6490
         |||   |||   |||||||||||||:::|||||||||||||||:::|
1264 oValThr...ValAsnThrValHisGluThrGlnGlyLysThrPheAspA 1280

6491 ATGTGGTATTGTTTAGGACGAAGAAAGCCGATGACTCCCTATTC...... 6534
     |||||::::::   |||   :::   ::::::::::::::::::
1280 spValIleValValArgLeuGluThrThrGluAsnGluIleTyrProGly 1296

6535 ......ACTAAACAACCGCATATACTTGTTGGTTTGTCGAGACACACACG 6578
                ||||||:::   :::|||:::   :::||||||   ||
1297 GlyArgLysGlyGlnProTyrGluIleValAlaThrThrArgHisArgAr 1313

6579 CTCACTGGTTTATGCCGCTCTGAGCTCAAAGTTGGACGATAAGGTCGGCA 6628
     ||||||||||||            :::   ::::::|||::::::
1313 gSerLeuValTyr.........TyrThrAlaIleGluAspArgLeuPheG 1327

6629 CATATATTAGCGACGCGTCACCTCAATCAGTATCCGACGCT...TTGCTT 6675
          |||||||||         |||:::|||   ::::::   |||:::
1327 luAspIleSerAspMet......GlnAspValMetGluSerLysLeuMet 1341

6676 CACACGTTCGCCCCG 6690
       ::::::::::|||
1342 LysSerLeuCysPro 1346
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 42.21 Seconds
(without alignments)
97.211 Million cell updates/sec

Title: US-09-301-906-5
Perfect score: 605
Sequence: 1 VSGSVSALRGDGKKVLMEAR.....EPAQSVPQQSLEEKAALTFG 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	605	100.0	120	21	Y58143	GLRaV-3 polyprotei
2	605	100.0	2237	21	Y58148	GLRaV-3 polyprotei
3	80.5	13.3	3010	15	R53417	Blood transmiscibl
4	78	12.9	483	14	R43893	elav. Drosophila
5	78	12.9	485	19	W60577	Drosophila neuron-
6	74	12.2	1301	18	W29115	FMR2P protein asso
7	71.5	11.8	3011	13	R22154	NANBV Hutch c59 is
8	71	11.7	285	19	W98314	H. pylori GHPO 894
9	70	11.6	1780	19	W53863	Human gravin polyp
10	69.5	11.5	380	12	R10748	Non-A non-B hepati
11	69.5	11.5	1061	20	W87504	Human N-methyl-D-a
12	69.5	11.5	1212	20	W87503	Human N-methyl-D-a
13	69.5	11.5	1219	18	W25763	Amino acid sequenc
14	69.5	11.5	3011	15	R66995	Hepatitis C virus
15	69	11.4	509	17	W05399	Human clone 65 pro
16	69	11.4	730	20	Y43385	S. pneumoniae PspC
17	69	11.4	1237	21	Y81609	Streptococcus pneu
18	68.5	11.3	248	18	W20433	H. pylori secreted
19	68.5	11.3	248	18	W20610	H. pylori secreted
20	68.5	11.3	449	19	W37129	Hepatitis C virus
21	68.5	11.3	1528	18	W33363	Human multidrug re
22	68.5	11.3	3011	14	R34468	Encoded by full-le
23	68	11.2	328	19	W36984	E. coli ZipA prote
24	68	11.2	390	17	W05283	Truncated human ta
25	68	11.2	441	15	R58810	Human tau protein.
26	68	11.2	441	17	W05282	Human tau protein.
27	68	11.2	441	18	W34856	Human tau protein.
28	68	11.2	441	21	Y81386	Human paired helic
29	68	11.2	3969	15	R52971	Product of the cDN
30	67.5	11.2	391	15	R50080	NANBH virus antige
31	67.5	11.2	1527	20	Y43543	A human MPR-relate
32	67.5	11.2	2783	13	R23962	AFP-1. Homo sapie
33	67.5	11.2	2783	13	R23963	AFP-1 (Ala 2460 Va
34	67	11.1	554	19	Y20763	Human neurofilamen
35	67	11.1	1400	15	R44514	MLL amino acid seq
36	67	11.1	1400	17	R92705	MLL cDNA clone 14-
37	67	11.1	3910	14	R38470	ALL-1 protein. Ho
38	67	11.1	3910	16	R66462	ALL-1 (acute lymph
39	66.5	11.0	238	18	W14587	Streptococcus pneu
40	66.5	11.0	250	20	Y41519	Fragment of human
41	66.5	11.0	346	21	Y58628	Protein regulating
42	66.5	11.0	411	17	R90934	HCV NS5 domain ant
43	66.5	11.0	479	20	Y03184	Topoisomerase II b
44	66.5	11.0	499	19	W67010	HCV non-structural
45	66.5	11.0	504	12	R14553	Chimpanzee HCV clo

ALIGNMENTS

RESULT 1
Y58143
ID Y58143 standard; Protein; 120 AA.
XX
AC Y58143;
XX
DT 07-MAR-2000 (first entry)
XX
DE GLRaV-3 polyprotein proteinase domain.
XX
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody.
XX
OS Grapevine leafroll-associated virus 3.
XX
PN WO9955880-A1.
XX
PD 04-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09307.
XX
PR 29-APR-1998; 98US-0083404.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Ling K;
XX
DR WPI; 2000-062035/05.
DR N-PSDB; Z49203.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful

PT for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 4; Fig 6; 84pp; English.
XX
CC This sequence represents the proteinase domain of the
CC polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 120 AA;

Query Match 100.0%; Score 605; DB 21; Length 120;
Best Local Similarity 100.0%; Pred. No. 3.7e-61;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKRE 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vsgsvsalrgdgkkvlmeartstsatsdvsdfdvvfeavsnallvvhyhrvvpyapvkre 60

Qy      61 QPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALTFG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qpkpavkqdeqkpkrqashwavkptavgvhvplpkkqealepaqsvpqqsleekaaltfg 120
```

RESULT 2
Y58148
ID Y58148 standard; Protein; 2237 AA.
XX
AC Y58148;
XX
DT 07-MAR-2000 (first entry)
XX
DE GLRaV-3 polyprotein.
XX
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT Domain 85..204
FT /note= "Proteinase domain (Y58143)"
FT Domain 460..731
FT /note= "Methyltransferase domain (Y58144)"
FT Domain 1922..2212
FT /note= "Helicase domain (Y58145)"
XX
PN WO9955880-A1.
XX
PD 04-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09307.
XX
PR 29-APR-1998; 98US-0083404.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Ling K;
XX
DR WPI; 2000-062035/05.
DR N-PSDB; Z49208.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
PT for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 3; Fig 16; 84pp; English.
XX
CC This sequence represents the polyprotein encoded by open reading
CC frame (ORF) 1a from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 2237 AA;

Query Match 100.0%; Score 605; DB 21; Length 2237;
Best Local Similarity 100.0%; Pred. No. 1.5e-59;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKRE 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 vsgsvsalrgdgkkvlmeartstsatsdvsdfdvvfeavsnallvvhyhrvvpyapvkre 144

Qy      61 QPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALTFG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     145 qpkpavkqdeqkpkrqashwavkptavgvhvplpkkqealepaqsvpqqsleekaaltfg 204
```

RESULT 3
R53417
ID R53417 standard; Protein; 3010 AA.
XX
AC R53417;
XX
DT 17-JAN-1995 (first entry)
XX
DE Blood transmiscible NANBHV protein.
XX
KW Polymerase chain reaction; PCR; amplify; primer; non-A, non-B hepatitis;
KW NANBH; virus; blood transmissible; detection; hepatitis virus; RT-PCR;
KW C100 antibody; HCV RNA; NS5 region.
XX
OS Non-A, non-B hepatitis virus.

```
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 222
FT                   /label= His, Arg
FT   Misc-difference 226
FT                   /label= Cys, Arg
FT   Misc-difference 246
FT                   /label= Leu, Phe
FT   Misc-difference 263
FT                   /label= Asp, Asn
FT   Misc-difference 291
FT                   /label= Phe, Ser
FT   Misc-difference 311
FT                   /label= Gly, Asp
FT   Misc-difference 398
FT                   /label= Ser, Arg, Gly
FT   Misc-difference 400
FT                   /label= Thr, Ala
FT   Misc-difference 405
FT                   /label= Gln, Pro, Leu
FT   Misc-difference 410
FT                   /label= Lys, Arg
FT   Misc-difference 418
FT                   /label= Gly, Asp
FT   Misc-difference 430
FT                   /label= Asn, Asp
FT   Misc-difference 438
FT                   /label= Phe, Leu
FT   Misc-difference 478
FT                   /label= Arg, Lys
FT   Misc-difference 759
FT                   /label= Leu, Val
FT   Misc-difference 1017
FT                   /label= Ser, Asn
FT   Misc-difference 1036
FT                   /label= Thr, Ala
FT   Misc-difference 1056
FT                   /label= Glu, Asp
FT   Misc-difference 1201
FT                   /label= Met, Thr
FT   Misc-difference 1205
FT                   /label= Met, Ile
FT   Misc-difference 1255
FT                   /label= Asn, Tyr
FT   Misc-difference 1263
FT                   /label= Gly, Asp
FT   Misc-difference 1455
FT                   /label= Asn, Asp
FT   Misc-difference 1828
FT                   /label= Ala, Thr
FT   Misc-difference 1895
FT                   /label= Gly, Arg
FT   Misc-difference 1896
FT                   /label= Gly, Ile
FT   Misc-difference 2143
FT                   /label= Glu, Val
FT   Misc-difference 2144
FT                   /label= Asp, Glu
FT   Misc-difference 2462
FT                   /label= Cys, Arg
FT   Misc-difference 2486
FT                   /label= Val, Met
FT   Misc-difference 2488
FT                   /label= Lys, Gln
FT   Misc-difference 2844
FT                   /label= Leu, Met
FT   Misc-difference 2862
FT                   /label= Leu, Gln
FT   Misc-difference 2917
FT                   /label= Arg, Leu
FT   Misc-difference 2968
FT                   /label= Ser, Gly
FT   Misc-difference 2989
FT                   /label= Cys, Arg
FT   Misc-difference 2990
FT                   /label= Tyr, Cys
XX
PN   JP06105690-A.
XX
PD   19-APR-1994.
XX
PF   10-MAR-1992;   92JP-0051885.
XX
PR   10-MAR-1992;   92JP-0051885.
XX
PA   (KAEN/) KAENNO K.
XX
DR   WPI; 1994-163130/20.
DR   N-PSDB; Q63499.
XX
PT   Blood-transmissible non-A non-B hepatitis virus DNA - used for
PT   detection of hepatitis virus
XX
PS   Claim 1; Page 8-20; 22pp; Japanese.
XX
CC   This sequence is encoded by the genome of a blood transmissible non-A,
CC   non-B hepatitis (NANBH) virus. The cDNA sequence was isolated using the
CC   primers given in Q63500-35.  The amplified fragments are used in the
CC   detection of hepatitis virus.  The target DNA was isolated from serum
CC   of chronically infected NANBH patients who were C100 antibody-positive
CC   and HCV RNA (NS5 region) positive.  Reverse transcription-PCR and PCR
CC   were performed on cDNA and the total human NANBH DNA was constructed
CC   from 23 clones.
XX
SQ   Sequence   3010 AA;


  Query Match             13.3%;  Score 80.5;  DB 15;  Length 3010;
  Best Local Similarity   21.1%;  Pred. No. 4.5;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       :  |  |:  |
Db    2237 mggnitrvesenkvvildsfdplraeederevsvaaeilrks------kkfppalpiwar 2290

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | : :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    2291 pdynpplleswkspdyvppavhgcplppttgppippprkkrtvvltestvssalaelatk 2350

Qy     118 TFG 120
           |||
Db    2351 tfg 2353


RESULT   4
R43893
ID   R43893 standard; protein; 483 AA.
XX
AC   R43893;
XX
DT   01-JUN-1994  (first entry)
XX
DE   elav.
XX
KW   Drosophila; neuron-specific protein; elav; K3; human; Hel-N1; c-fos;
KW   development; central nervous system; hypotrophy; immune cell; c-myb;
KW   embryonic lethality; primer; RNP-1; RNA recognition motif; RRM; CNS;
KW   oncoprotein; lymphokine; c-myc; GM-CSF; interferon; differentiation;
KW   interleukin; modulation; expression; 3'-untranslated region; cancer;
KW   mRNA instability sequence; eukaryotic cell; regulation; diagnosis;
KW   cell proliferation; mammal; 3'-UTR; transgenic animal; ss.
XX
OS   Drosophila melanogaster.
XX
FH   Key             Location/Qualifiers
FT   Domain          151..156
```

```
FT                   /note= "RRM1 RNP2"
FT   Domain          203..210
FT                   /note= "RRM1 RNP1"
FT   Domain          250..255
FT                   /note= "RRM2 RNP2"
FT   Domain          290..297
FT                   /note= "RRM2 RNP1"
FT   Domain          404..409
FT                   /note= "RRM3 RNP2"
FT   Domain          443..450
FT                   /note= "RRM3 RNP1"
XX
PN   WO9323531-A.
XX
PD   25-NOV-1993.
XX
PF   10-MAY-1993;   93WO-US04194.
XX
PR   11-MAY-1992;   92US-0881075.
XX
PA   (UYDU-) UNIV DUKE.
XX
PI   Keene JD,  King PH,  Levine T;
XX
DR   WPI; 1993-386565/48.
XX
PT   Polypeptide(s) Hel-N1, elav, K3 and HuD and their fragments, DNA
PT   and RNA sequences - bind to instability sequences of mRNA
PT   encoding onco:proteins and lymphokine(s) and modulate expression,
PT   used to regulate cell proliferation, as genetic or neuro-genetic
PT   disease markers, etc.
XX
PS   Claim 2; Fig 1; 60pp; English.
XX
CC   The sequences given in R43893-95 are the two Drosophila neuron-
CC   specific proteins, elav and K3, and the human counterpart, Hel-N1.
CC   Elav is known to be involved in the early development of the central
CC   nervous system (CNS).  Homozygous mutations of this gene locus give
CC   rise to numerous structural defects and hypotrophy of the CNS leading
CC   to embryonic lethality.  Degenerate primers designed to simulate the
CC   RNP-1 octamer sequence present in two of the three RNA recognition
CC   motifs (RRMs) of Drosophila elav, thereby isolating cDNA encoding a
CC   neuron-specific protein, Hel-N1.  Elav, K3 and Hel-N1 bind to mRNAs
CC   encoding oncoproteins or lymphokines, pref. c-myc, c-fos, c-myb,
CC   GM-CSF, an interferon or an interleukin, and can provide, in cell
CC   cultures, or in vivo, modulation of their expression.  They bind to
CC   the 3'-untranslated regions of mRNA, particularly mRNA instability
CC   sequences in eukaryotic cells.  These proteins and the DNAs encoding
CC   them may be used in the regulation of cell proliferation, particularly
CC   to take cells out of the proliferative state into a state of
CC   differentiation.  They are used for regenerating tissue, especially
CC   neuronal, the regulation of immune cell gene expression and
CC   stimulating or supressing mammalian cell growth.  Transgenic animals
CC   encoding these proteins can be used in pharmaceutical tests for cancer,
CC   immune function and/or neurological growth.  The proteins may also be
CC   used for in vivo targetting of substances, diagnosing pathology
CC   specimens if neuronal origin, and as genetic or neurogenetic disease
CC   markers in diagnosis and therapy.
XX
SQ   Sequence   483 AA;


  Query Match             12.9%;  Score 78;  DB 14;  Length 483;
  Best Local Similarity   25.9%;  Pred. No. 0.86;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
           | |     :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      11 gggvdtqaqlmqsaaaaaavaatnaaaapvqnaa------avaaaaqlqqqqvqqailqv 64

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      65 qqqqtqq----avaaaaaavtqqlqqqqqavvaqqavvqqqqqqaaav 108
```

```
RESULT   5
W60577
ID   W60577 standard; protein; 485 AA.
XX
AC   W60577;
XX
DT   07-SEP-1998  (first entry)
XX
DE   Drosophila neuron-specific protein elav.
XX
KW   Neuron-specific protein; growth regulatory protein; oncoprotein;
KW   proto-oncogene; cytokine; lymphokine; chromosome mapping; drosophila;
KW   human; functionally related protein; elav.
XX
OS   Drosophila sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 207
FT                   /note= "unknown or unidentified"
FT   Misc-difference 211
FT                   /note= "unknown or unidentified"
XX
PN   US5773246-A.
XX
PD   30-JUN-1998.
XX
PF   07-JUN-1995;   95US-0478675.
XX
PR   15-SEP-1993;   93US-0120827.
PR   11-MAY-1992;   92US-0881075.
PR   07-JUN-1995;   95US-0478675.
XX
PA   (GAOF/) GAO F.
PA   (KEEN/) KEENE J D.
PA   (LEVI/) LEVINE T.
XX
PI   Gao F,  Keene JD,  Levine T;
XX
DR   WPI; 1998-387003/33.
XX
PT   Use of proteins which bind RNA - for obtaining a cDNA library
PT   containing members encoding structurally or functionally related
PT   proteins from total cell mRNA.
XX
PS   Disclosure; Fig 1A-C; 67pp; English.
XX
CC   This represents the drosophila neuron-specific protein elav. The
CC   invention provides a method for obtaining a cDNA library having members
CC   encoding a group of structurally or functionally related proteins from
CC   total cell mRNA. The method comprises binding RNA representing total cell
CC   mRNA to a protein that binds RNA and has specific binding to untranslated
CC   regions of a subset of the total cell mRNA, where the protein that binds
CC   RNA has been purified to remove any other protein that binds RNA. The
CC   resulting bound products are separated and a cDNA library is prepared.
CC   The protein that binds RNA may be Hel-N1, Hel-N2, CArG, DT-7, K1, K2,
CC   K3, HuD, HuC, elav, rbp9, eIF4B, sxl, tra-2, AUBF, AUF, ASF/SF2, U2AF,
CC   SC35, or hnRNP proteins. The cDNAs obtained can be used for chromosome
CC   mapping and genome sequencing of the structurally or functionally related
CC   genes encoding growth regulatory proteins, proto-oncogenes, cytokines,
CC   lymphokines, or anti-oncogene proteins. They can also be used in
CC   diagnostic methods for determining imbalances in such genes and the
CC   efficacy of various treatments to correct any imbalance.
XX
SQ   Sequence   485 AA;


  Query Match             12.9%;  Score 78;  DB 19;  Length 485;
  Best Local Similarity   25.9%;  Pred. No. 0.86;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
```

```
          | |     :   | :| : |:  :     | ||        |   | ::::| : |: |
Db     13 gggvdtqaqlmqsaaaaaavaatnaaaapvqnaa------avaaaaqlqqqqvqqailqv 66

Qy     70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
          :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db     67 qqqqtqq----avaaaaaavtqqlqqqqqavvaqqavvqqqqqqaaav 110


RESULT   6
W29115
ID   W29115 standard; Protein; 1301 AA.
XX
AC   W29115;
XX
DT   15-JAN-1998  (first entry)
XX
DE   FMR2P protein associated with FRAXE mental retardation.
XX
KW   FMR2 gene; FMR2P; FRAXE; rate folate-sensitive fragile site;
KW   X-linked mental retardation; diagnosis; therapy; antibody.
XX
OS   Homo sapiens.
XX
PN   WO9723610-A1.
XX
PD   03-JUL-1997.
XX
PF   20-DEC-1996;   96WO-AU00825.
XX
PR   22-DEC-1995;   95AU-0007366.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Gecz J,  Mulley JC;
XX
DR   WPI; 1997-351051/32.
DR   N-PSDB; T85728.
XX
PT   DNA containing gene associated with FRAXE mental retardation -
PT   useful for diagnosis and therapy of FRAXE mental retardation
XX
PS   Claim 36; Fig 2; 39pp; English.
XX
CC   This polypeptide comprises FMR2P protein encoded by the human FMR2
CC   gene (see T85728).  Absence of FMR2P appears to be the molecular
CC   basis for FRAXE (a rare folate-snesitive fragile site), a
CC   non-specific, mild, X-linked form of mental retardation.  This can
CC   be caused either by CCG expansion within the 5'-untranslated region
CC   of the FMR2 gene (see T85728) or by deletion of coding sequences.
CC   An isolated FMR2 cDNA sequence can be used in a claimed recombinant
CC   expression system to produce a protein product in transformed host
CC   cells.  This can be used to raise polyclonal or monoclonal
CC   antibodies that form the basis of an immunocytochemical test for
CC   the presence of FMR2P.  The protein encoded by the normal
CC   chromosome can also be used to treat FRAXE mental retardation by
CC   replacing, repairing or compensating for the FRAXE allele.
XX
SQ   Sequence   1301 AA;


  Query Match             12.2%;  Score 74;  DB 18;  Length 1301;
  Best Local Similarity   32.4%;  Pred. No. 8.5;
  Matches   23;  Conservative   12;  Mismatches   28;  Indels    8;  Gaps    3;

Qy     50 RVVPYAPVKREQPKPAVKQDEQKPKRQASHWAV-KPTAVGVHVPLPKKQEALEPAQSVPQ 108
          ||  :: :     ||  |:   ||||| :  || ||   |      :| : :| |: :|:
Db    802 rvpghsslhaapakpdhketatkpkrqtavtavekpapkg-----krkhkpievaekipe 856

Qy    109 --QSLEEKAAL 117
            | |||   :
Db    857 kkqrleeatti 867
```

```
RESULT   7
R22154
ID   R22154 standard; Protein; 3011 AA.
XX
AC   R22154;
XX
DT   07-JUL-1992  (first entry)
XX
DE   NANBV Hutch c59 isolate genome.
XX
KW   Hepatitis C virus; non-A non-B virus; HCV-Hc59; antibodies;
KW   vaccine; assay; detection.
XX
OS   NANBV Hutch c59 isolate.
XX
FH   Key             Location/Qualifiers
FT   Region          1..326
FT                   /label= NANBV_structural_proteins
FT                   /note= "Seq No 1 (R22137)"
FT   Misc-difference 3
FT                   /label= Asn
FT                   /note= "or Ile according to Seq No 1 (R22137), see CC"
FT   Misc-difference 321
FT                   /label= Asn
FT                   /note= "or Asp according to Seq No 1 (R22137), see CC"
FT   Region          1..120
FT                   /label= capsid
FT   Region          121..326
FT                   /label= envelope_protein
FT   Peptide         1..20
FT                   /label= pref._capsid_antigen
FT   Peptide         21..40
FT                   /label= pref._capsid_antigen
FT   Peptide         2..40
FT                   /label= pref._capsid_antigen
FT   Peptide         1..74
FT                   /label= pref._capsid_antigen
FT   Peptide         69..120
FT                   /label= pref._capsid_antigen
FT   Peptide         121..176
FT                   /label= pref._envelope_antigen
FT   Region          386..411
FT                   /label= V_variable_region
FT   Region          246..275
FT                   /label= V1_variable_region
FT   Region          456..482
FT                   /label= V2_variable_region
FT   Region          2356..2379
FT                   /label= V3_variable_region
FT   Peptide         391..404
FT                   /label= pref._V_fragment
FT   Peptide         246..256
FT                   /label= pref._V1_fragment
FT   Peptide         461..466
FT                   /label= pref._V2_fragment
FT   Peptide         473..482
FT                   /label= pref._V2_fragment
XX
PN   WO9203458-A.
XX
PD   05-MAR-1992.
XX
PF   23-AUG-1991;   91WO-US06037.
XX
PR   21-NOV-1990;   90US-0616369.
PR   25-AUG-1990;   90US-0573643.
XX
PA   (NYBL-) NEW YORK BLOO DCENT.
PA   (PHAA ) PHARMACIA GENETIC ENG INC.
XX
PI   Zebedee S,  Inchauspe G,  Nasofe MS,  Prince AM;
XX
DR   WPI; 1992-096821/12.
```

```
DR   N-PSDB; Q22871.
XX
PT   Deoxyribonucleic acid sequence encoding non-A, non-B hepatitis
PT   virus - obtd. Hutch C59 subgroup encoding polypeptide(s), useful
PT   as vaccines, and immuno reactive Abs for diagnosis of virus
XX
PS   Disclosure; Page 159-173; 225pp; English.
XX
CC   One Hutch strain (HCV-H) of NANBV, designated the Hutch c59 isolate
CC   (HCV-Hc59) was propagated through passage in animals and the
CC   entire viral genome was cloned and sequenced.
CC   The sequence in R22137 comprises two amino acids, indicated in
CC   the features, which differ from the sequence represented here.
CC   The proteins and peptides (see features) and antibodies against
CC   them are useful for the prepn. of vaccines and inoculums against
CC   NANBV and in immunological assays for detection of viral infection.
CC   The indicated ranges of amino acids correspond to portions of the
CC   V, V1, V2 and V3 regions that contain the  greatest amt. of
CC   diversity when compared to known HCV isolates.  The polynucleotides
CC   encoding these may be used as primers, probes, or a nucleic acids.
XX
SQ   Sequence   3011 AA;


  Query Match             11.8%;  Score 71.5;  DB 13;  Length 3011;
  Best Local Similarity   22.0%;  Pred. No. 47;
  Matches   27;  Conservative   23;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       |  |  ||  |
Db    2237 mggnitrvesenkvvildsfdplvaeederevsvpaeilrks------rrfapalpvwar 2290

Qy      60 EQPKPAVKQDEQKPKRQ--ASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | : :  :||  :    |    |      || |:|:  :   :|    :| | |
Db    2291 pdynpllvetwkkpdyeppvvhgcplpppprsppvpppprkkrtvvltestlptalaelatk 2350

Qy     118 TFG 120
           :||
Db    2351 sfg 2353


RESULT   8
W98314
ID   W98314 standard; Protein; 285 AA.
XX
AC   W98314;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 894 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; X14033.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 413-415; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   285 AA;


  Query Match             11.7%;  Score 71;  DB 19;  Length 285;
  Best Local Similarity   25.5%;  Pred. No. 2.7;
  Matches   28;  Conservative   18;  Mismatches   48;  Indels   16;  Gaps    4;

Qy       8 LRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVK 67
             || |  : | :| | |  |  ::  :      :||           :   :|:||   |
Db      35 lredapeplaqagt-tkvtmslasintnsntktna----------esakpkeepkekpk 82

Qy      68 QDEQK---PKRQASHWAVKPTAVGVHVPLPKKQEALEP-AQSVPQQSLEE 113
           ::| |   ||:: :    ||       | || :   ||  :  |:  :||
Db      83 keepkkeepkkevtkpkpkpkpkpkpkpkpepkpepkpepkpepkvee 132


RESULT   9
W53863
ID   W53863 standard; peptide; 1780 AA.
XX
AC   W53863;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Human gravin polypeptide.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS   Homo sapiens.
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
DR   N-PSDB; V23545.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 1; Column 19-32; 32pp; English.
XX
CC   This sequence corresponds to the human gravin polypeptide, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
```

```
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence   1780 AA;


  Query Match             11.6%;  Score 70;  DB 19;  Length 1780;
  Best Local Similarity   29.3%;  Pred. No. 36;
  Matches   17;  Conservative   10;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLE 112
           |  |  :|| : :: |:  ||: ||  : | |         |:|  |  :  | :| |
Db     224 aaskesepkqstekpeetlkreqshaeisppaesgqaveeckeegeekqekepsksae 281


RESULT  10
R10748
ID   R10748 standard; Protein; 380 AA.
XX
AC   R10748;
XX
DT   22-APR-1991  (first entry)
XX
DE   Non-A non-B hepatitis specific antigenic protein encoded by phage
DE   clone lambda HC2232.
XX
KW   Non-A non-B hepatitis; antigenic protein; NANBH; phage clone;
KW   immunoassay; antibodies; diagnosis.
XX
OS   Homo sapiens.
XX
PN   WO9101376-A.
XX
PD   07-FEB-1991.
XX
PF   13-JUL-1990;   90WO-JP00906.
XX
PR   19-OCT-1989;   89JP-0270398.
PR   14-JUL-1989;   89JP-0182073.
PR   19-JUL-1989;   89JP-0184739.
PR   22-JUL-1989;   89JP-0189874.
PR   27-JUL-1989;   89JP-0192721.
PR   29-JUL-1989;   89JP-0195413.
PR   03-AUG-1989;   89JP-0200217.
PR   10-AUG-1989;   89JP-0205722.
PR   21-SEP-1989;   89JP-0243304.
PR   22-SEP-1989;   89JP-0245268.
XX
PA   (CHUS ) CHUGAI SEIYAKU KK.
XX
PI   Arima T,  Yamamoto O,  Tsuchiya M,  Oshima M;
XX
DR   WPI; 1991-058149/08.
DR   N-PSDB; Q10529.
XX
PT   Antigenic protein specific for non-A, non-B hepatitis - and cDNA
PT   coding for it which corresponds to RNA of infected liver tissue
PT   or serum
XX
PS   Disclosure; Fig 9; 69pp; Japanese.
XX
CC   The DNA corresponding to an RNA isolated directly from infected
CC   human liver tissue or serum is used to synthesise cDNA.  A library
CC   is constructed using a lambda gt11 cloning system.  This library
CC   is screened and cloned for antigenic activity.  Clones isolated
CC   can be incorporated into a vector plasmid, which is then inserted
CC   into E.coli to give a transformant which expressed an antigenic
CC   protein having this amino acid sequence.
CC   The product may be used for the immunoassay of antibodies to NANBH
CC   antigen in samples of serum etc. for diagnostic purposes.
CC   See also Q10523-540.
XX
SQ   Sequence   380 AA;


  Query Match             11.5%;  Score 69.5;  DB 12;  Length 380;
  Best Local Similarity   22.6%;  Pred. No. 5.8;
  Matches   28;  Conservative   22;  Mismatches   63;  Indels   11;  Gaps    4;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYA-PV-K 58
           : |::: :  : | |::::     |  |  :  |  |       ::   |  | | |:
Db     118 mggnitrvesenkvvildsfeplraeederevsvaae-------ilrktrrfpaampiwa 170

Qy      59 REQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAA 116
           |    | : |  : |       |    | |    || |:::  :   :|    :| | |
Db     171 rpdynppllqswkdpdyvppvvhgcplppakappvppprrkrtvvltestvssalaelat 230

Qy     117 LTFG 120
           |||
Db     231 ktfg 234


RESULT  11
W87504
ID   W87504 standard; Protein; 1061 AA.
XX
AC   W87504;
XX
DT   23-FEB-1999  (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA24.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW   NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS   Homo sapiens.
XX
PN   US5849895-A.
XX
PD   15-DEC-1998.
XX
PF   20-APR-1994;   94US-0231193.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI   Daggett LP,  Lu C;
XX
DR   WPI; 1999-069812/06.
DR   N-PSDB; V82910.
XX
PT   DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT   assembly of functional glutamate receptor subunits
XX
PS   Example 3; Columns 261-268; 203pp; English.
XX
CC   The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC   receptor subunit (NMDAR). The nucleic acid sequence does not contain the
CC   860 5'-most nucleotides, has an additional 11 nucleotides (V82891)
CC   between nucleotides 1300 and 1301, an additional 24 nucleotides (V82890)
CC   inserted between nucleotides 23050 and 2351, as set forth in V82889. The
CC   cDNA sequence is derived from clone NMDA24. The NMDAR subunits contribute
CC   to the formation of NMDA-activated cation-selective ion channels. In
CC   addition to being useful for the production of NMDA receptor subunit
CC   proteins, the nucleic acids are also useful as probes to identify and
CC   isolate nucleic acids encoding related receptor subunits. Functional
CC   glutamate receptors can be assembled from several NMDA receptor subunit
CC   proteins of one type (homomeric) or from combinations of subunit proteins
CC   of different types (heteromeric). The present invention also comprises
CC   methods for using such receptor subunits to identify and characterise
```

```
CC   compounds which affect the function of such receptors , e.g. agonists,
CC   antagonists and modulators of glutamate receptor function. The invention
CC   also comprises methods for determining whether unknown protein(s) are
CC   functional as NMDA receptor subunits.
XX
SQ   Sequence   1061 AA;


  Query Match             11.5%;  Score 69.5;  DB 20;  Length 1061;
  Best Local Similarity   27.5%;  Pred. No. 21;
  Matches   22;  Conservative   13;  Mismatches   32;  Indels   13;  Gaps    4;

Qy      49 HRVVPYAPVKREQPKPA----VKQDEQKPKRQASHWA--VKPTAVGV------HVPLPKK 96
           ||  |    :  :|:||     :|    |:| |: |   : |: :||      | | | :
Db     650 hrgrpartsrparprpacsrfcrqpatwprraatpwtaplapsrigvaaavrphrp-par 708

Qy      97 QEALEPAQSVPQQSLEEKAA 116
             | || : |  :  :  |
Db     709 prglapahacppptrpqsra 728


RESULT  12
W87503
ID   W87503 standard; Protein; 1212 AA.
XX
AC   W87503;
XX
DT   23-FEB-1999  (first entry)
XX
DE   Human N-methyl-D-aspartate receptor subunit encoded by clone NMDA22.
XX
KW   Human; N-methyl-D-aspartate receptor; NMDAR2C;
KW   NMDA-activated cation-selective ion channel; glutamate receptor.
XX
OS   Homo sapiens.
XX
PN   US5849895-A.
XX
PD   15-DEC-1998.
XX
PF   20-APR-1994;   94US-0231193.
XX
PR   20-APR-1994;   94US-0231193.
PR   20-APR-1993;   93US-0052449.
XX
PA   (SIBI-) SIBIA NEUROSCIENCES INC.
XX
PI   Daggett LP,  Lu C;
XX
DR   WPI; 1999-069812/06.
DR   N-PSDB; V82909.
XX
PT   DNA encoding N-methyl-D-aspartate receptor subunit - useful for the
PT   assembly of functional glutamate receptor subunits
XX
PS   Example 3; Columns 253-262; 203pp; English.
XX
CC   The present sequence represents a human N-methyl-D-aspartate (NMDA)
CC   receptor subunit (NMDAR). The nucleic acid sequence does not contain
CC   the 366 5'-most nucleotides, by the insertion of 11 nucleotides between
CC   nucleotides 1300 and 1301, nor the 15 nucleotides at positions 1960-1974,
CC   nor the 1061 3' nucleotides, as set forth in V82889. The cDNA sequence is
CC   derived from clone NMDA21. The NMDAR subunits contribute to the formation
CC   of NMDA-activated cation-selective ion channels. In addition to being
CC   useful for the production of NMDA receptor subunit proteins, the nucleic
CC   acids are also useful as probes to identify and isolate nucleic acids
CC   encoding related receptor subunits. Functional glutamate receptors can be
CC   assembled from several NMDA receptor subunit proteins of one type
CC   (homomeric) or from combinations of subunit proteins of different
CC   types (heteromeric). The present invention also comprises methods
CC   for using such receptor subunits to identify and characterise
CC   compounds which affect the function of such receptors , e.g. agonists,
CC   antagonists and modulators of glutamate receptor function. The invention
CC   also comprises methods for determining whether unknown protein(s) are
CC   functional as NMDA receptor subunits.
XX
SQ   Sequence   1212 AA;


  Query Match             11.5%;  Score 69.5;  DB 20;  Length 1212;
  Best Local Similarity   27.5%;  Pred. No. 25;
  Matches   22;  Conservative   13;  Mismatches   32;  Indels   13;  Gaps    4;

Qy      49 HRVVPYAPVKREQPKPA----VKQDEQKPKRQASHWA--VKPTAVGV------HVPLPKK 96
           ||  |    :  :|:||     :|    |:| |: |   : |: :||      | | | :
Db     801 hrgrpartsrparprpacsrfcrqpatwprraatpwtaplapsrigvaaavrphrp-par 859

Qy      97 QEALEPAQSVPQQSLEEKAA 116
             | || : |  :  :  |
Db     860 prglapahacppptrpqsra 879


RESULT  13
W25763
ID   W25763 standard; Protein; 1219 AA.
XX
AC   W25763;
XX
DT   03-DEC-1997  (first entry)
XX
DE   Amino acid sequence of pCaR/R1.
XX
KW   Chimeric receptor; extracellular domain; seven transmembrane domain;
KW   intracellular cytoplasmic tail domain; metabotropic glutamate receptor;
KW   MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;
KW   neurological disease.
XX
OS   Chimeric - Homo sapiens.
XX
PN   WO9705252-A2.
XX
PD   13-FEB-1997.
XX
PF   25-JUL-1996;   96WO-US12336.
XX
PR   26-JUL-1995;   95US-0001526.
XX
PA   (NPSP-) NPS PHARM INC.
XX
PI   Fuller FH,  Hammerland LG,  Krapcho KJ;
XX
DR   WPI; 1997-145690/13.
DR   N-PSDB; T86166.
XX
PT   Chimeric receptors comprising metabotropic glutamate receptor and
PT   calcium receptor - used for screening for neurologically active
PT   compounds
XX
PS   Example 3; Fig 3; 177pp; English.
XX
CC   This sequence represents a chimeric receptor.  The chimeric receptor
CC   comprises an extracellular domain, a seven transmembrane domain, and
CC   an intracellular cytoplasmic tail domain, and a sequence of at least
CC   6 contiguous amino acids is homologous to a sequence of a metabotropic
CC   glutamate receptor (MGR), and a sequence of at least 6 contiguous amino
CC   acids is homologous to a sequence of a calcium receptor (CR).  The
CC   chimeric receptor may be used for screening for compounds that bind to
CC   or modulate the activity of MGR or CR and for determining the site-of-
CC   action of a CR active compound. The compounds can be used in the
CC   treatment of neurological diseases and disorders.  They can also be
CC   used as diagnostic agents.  Chimeric receptors such as this, allow the
CC   coupling of certain functional aspects of an MGR with certain functional
CC   aspects of a CR.  They allow for more efficient high-throughput screening
CC   of compounds.
XX
SQ   Sequence   1219 AA;
```

```
  Query Match             11.5%;  Score 69.5;  DB 18;  Length 1219;
  Best Local Similarity   32.4%;  Pred. No. 25;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy       9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
           :| ||  |  :  ||    :| : |      |   |:  :||  ||  |  ||    | |
Db     958 qgsgks-ltfsdastktlynveeedntpsahfsppsspsmvv--hrrgp--pvattpplp 1012

Qy      65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| |   |:  |||::|    | |  |||
Db    1013 phltaeetplflad--svipk--glppplpqqqpqqpppqqppqq 1053


RESULT  14
R66995
ID   R66995 standard; Protein; 3011 AA.
XX
AC   R66995;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Hepatitis C virus gene HC-J1/protein.
XX
KW   Hepatitis C virus; HCV gene HC-J1/protein; specific antibodies.
XX
OS   Hepatitis c virus.
XX
PN   JP06284887-A.
XX
PD   11-OCT-1994.
XX
PF   10-DEC-1993;   93JP-0345753.
XX
PR   10-DEC-1992;   92JP-0360705.
XX
PA   (IMMO ) IMMUNO JAPAN KK.
XX
DR   WPI; 1994-362594/45.
DR   N-PSDB; Q74770.
XX
PT   HCV genes and the corresponding proteins - used in the production
PT   of anti-HCV antibodies and the detection of HCV infection
XX
PS   Claim 11; Pages 18-32; 35pp; Japanese.
XX
CC   Q74770 encodes R66995 the HC-J1/protein, the cDNA can be used in
CC   the construction of an expression vector for the transformation
CC   of a host cell. The host cell can then be used in the production
CC   of proteins and peptides, useful in the preparation of monoclonal
CC   and polyclonal HCV-specific antibodies.
XX
SQ   Sequence   3011 AA;


  Query Match             11.5%;  Score 69.5;  DB 15;  Length 3011;
  Best Local Similarity   21.1%;  Pred. No. 79;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       |     |:  |
Db    2237 mggnitrvesenkvvildsfdplvaeedereisvpaeilrks------rrftqalpiwar 2290

Qy      60 EQPKPAVKQDEQKPKRQ--ASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | : :  :||  :    |    |      || |:|:  :   :|    :| | ||
Db    2291 pdynpplietwkkpnyeppvvhgcplpppqsppvppprkkrtvvltestlstalaelaak 2350

Qy     118 TFG 120
           :||
Db    2351 sfg 2353


RESULT  15
W05399
ID   W05399 standard; Protein; 509 AA.
XX
AC   W05399;
XX
DT   19-FEB-1998  (first entry)
XX
DE   Human clone 65 protein.
XX
KW   Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
KW   cellular signalling element; cellular structural element; malignancy;
KW   protein identification; functional domain; protein screening;
KW   cellular signal transduction process.
XX
OS   Homo sapiens.
XX
PN   WO9631625-A1.
XX
PD   10-OCT-1996.
XX
PF   04-APR-1996;   96WO-US04454.
XX
PR   03-APR-1996;   96US-0630915.
PR   07-APR-1995;   95US-0417872.
XX
PA   (CYTO-) CYTOGEN CORP.
PA   (UYNC-) UNIV NORTH CAROLINA.
XX
PI   Fowlkes DM,  Hoffman N,  Kay BK,  Mcconnell SJ,  Sparks AB;
XX
DR   WPI; 1996-465045/46.
DR   N-PSDB; T39799.
XX
PT   Identifying polypeptide(s) having specific functional domain (esp.
PT   SH3 domain) - comprises detecting selective binding to recognition
PT   unit, regardless of sequence homology
XX
PS   Claim 54; Fig 59; 174pp; English.
XX
CC   W05386-W05403 represent novel human and mouse Src-homology region 3 (SH3)
CC   domain containing proteins that can be used in the method of the
CC   invention. SH3 domain containing proteins play a role in signalling and
CC   structural elements of cells. The method of the invention is for
CC   identifying polypeptides containing functional domains of interest
CC   (especially SH3 domains). The method comprises contacting a multivalent
CC   recognition unit (RU) complex with a number of peptides and identifying
CC   polypeptides having a selective binding affinity for the RU complex. The
CC   method is based on functional similarities and does not rely on sequence
CC   similarities. Prior methods only gave limited success for identifying
CC   proteins which contain an SH3 domain due to the minimal sequence homology
CC   among known SH3 proteins. It has been found that small peptide RUs in
CC   multivalent form have reduced specificity for a given functional domain
CC   compared to monomer RUs. Multivalent RU complexes are particularly suited
CC   to screening for polypeptides containing functional domains that are
CC   similar to, but not identical in sequence to, the original target
CC   functional domain. The new method enables proteins having a common
CC   function to be identified. Identification of novel SH3 proteins will be
CC   useful for a better understanding of cell growth, malignancy, signal
CC   transduction processes, etc. New candidate drugs can be identified, and
CC   their specificities (e.g. pharmacological activities) can be assessed
CC   using the method of the invention.
XX
SQ   Sequence   509 AA;


  Query Match             11.4%;  Score 69;  DB 17;  Length 509;
  Best Local Similarity   29.0%;  Pred. No. 9.5;
  Matches   40;  Conservative   14;  Mismatches   42;  Indels   42;  Gaps    8;

Qy       3 GSVSALRGDGKKV---LMEARTSTSATSD------VSDFDVVFEAVSNALL-VVH----- 47
           |||| : | |: |     :|  | :|  |        |   | |   |     ||
Db     203 gsvspihgqgqvvenlkaqalcswtakkdnhlnfskhdiitvleqqenwwfgevhggrgw 262
```

```
Qy      48 ----YHRVVPYAPVKREQPKPAVKQDEQKPKRQASHWAV--KPTAVGVHVPLPKKQEALE 101
           |  :::| : ||||:|             :| : ||  |||:    |   ::  ||
Db     263 fpksyvkiipgsevkreep------------ealyaavnkkptsaaysv--geeyialy 307

Qy     102 PAQSVPQQSLEEKAALTF 119
           |  ||      |   |||
Db     308 pyssv------epgdltf 319
```

Search completed: January 31, 2001, 15:28:18
Job time: 147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 27.64 Seconds (without alignments) 77.961 Million cell updates/sec

Title: US-09-301-906-5
Perfect score: 605
Sequence: 1 VSGSVSALRGDGKKVLMEAR.....EPAQSVPQQSLEEKAALTFG 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	12.9	485	1	US-07-881-075-1	Sequence 1, Appli
2	78	12.9	485	1	US-08-120-827-1	Sequence 1, Appli
3	78	12.9	485	1	US-08-478-675-1	Sequence 1, Appli
4	70	11.6	1780	1	US-08-769-309A-5	Sequence 5, Appli
5	70	11.6	1780	3	US-08-994-570-5	Sequence 5, Appli
6	69.5	11.5	1199	1	US-08-041-538-2	Sequence 2, Appli
7	69.5	11.5	1199	1	US-08-463-642-2	Sequence 2, Appli
8	69.5	11.5	1199	1	US-08-455-602-2	Sequence 2, Appli
9	69.5	11.5	1199	2	US-08-465-157-2	Sequence 2, Appli
10	69.5	11.5	1199	4	PCT-US91-09422-2	Sequence 2, Appli
11	69.5	11.5	1219	2	US-08-687-289A-6	Sequence 6, Appli
12	68	11.2	328	2	US-08-651-818A-2	Sequence 2, Appli
13	67	11.1	1400	1	US-08-080-255-7	Sequence 7, Appli
14	67	11.1	1400	3	US-08-465-713-7	Sequence 7, Appli
15	67	11.1	1400	4	PCT-US93-05857-7	Sequence 7, Appli
16	66.5	11.0	504	1	US-07-853-985A-6	Sequence 6, Appli
17	66.5	11.0	504	1	US-07-681-703B-6	Sequence 6, Appli
18	66.5	11.0	504	1	US-08-184-236-6	Sequence 6, Appli
19	66.5	11.0	504	2	US-08-407-410B-6	Sequence 6, Appli
20	66.5	11.0	504	2	US-08-485-500-6	Sequence 6, Appli
21	66.5	11.0	504	4	PCT-US91-02370-6	Sequence 6, Appli
22	66.5	11.0	504	4	PCT-US94-04174-6	Sequence 6, Appli
23	66.5	11.0	541	4	PCT-US92-06965A-13	Sequence 13, Appl
24	66.5	11.0	872	1	US-08-491-357-3	Sequence 3, Appli
25	66.5	11.0	872	3	US-08-968-633-3	Sequence 3, Appli
26	66.5	11.0	872	3	US-09-196-466-3	Sequence 3, Appli
27	66.5	11.0	872	4	PCT-US96-10823-3	Sequence 3, Appli
28	66.5	11.0	1786	3	US-08-444-818-54	Sequence 54, Appl
29	66.5	11.0	2261	3	US-08-444-818-66	Sequence 66, Appl
30	66.5	11.0	2436	3	US-08-444-818-75	Sequence 75, Appl
31	66.5	11.0	2772	3	US-08-444-818-89	Sequence 89, Appl
32	66.5	11.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
33	66.5	11.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
34	66.5	11.0	2894	3	US-08-467-902A-23	Sequence 23, Appl
35	66.5	11.0	2955	2	US-08-443-260-3	Sequence 3, Appli
36	66.5	11.0	2955	3	US-08-442-805A-3	Sequence 3, Appli
37	66.5	11.0	2955	3	US-08-443-900A-3	Sequence 3, Appli
38	66.5	11.0	2955	3	US-08-444-818-124	Sequence 124, App
39	66.5	11.0	2995	3	US-08-444-818-138	Sequence 138, App
40	66.5	11.0	3011	1	US-08-440-103-36	Sequence 36, Appl
41	66.5	11.0	3011	1	US-08-440-542-36	Sequence 36, Appl
42	66.5	11.0	3011	1	US-07-910-760-10	Sequence 10, Appl
43	66.5	11.0	3011	1	US-08-440-519-10	Sequence 10, Appl
44	66.5	11.0	3011	1	US-08-231-368-36	Sequence 36, Appl
45	66.5	11.0	3011	1	US-08-440-210-36	Sequence 36, Appl

ALIGNMENTS

```
RESULT   1
US-07-881-075-1
; Sequence 1, Application US/07881075
; Patent No. 5444149
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  51
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/881,075
;      FILING DATE:  19920511
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5444149man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-154-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)521-4500
;      TELEFAX:  (703)486-2347
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-881-075-1


  Query Match           12.9%;  Score 78;  DB 1;  Length 485;
  Best Local Similarity  25.9%;  Pred. No. 0.18;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
```

```
          | |     :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      13 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 66

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      67 QQQQTQQ----AVAAAAAAVTQQLQQQQQAVVAQQAVVQQQQQQAAAV 110


RESULT   2
US-08-120-827-1
; Sequence 1, Application US/08120827
; Patent No. 5525495
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/120,827
;      FILING DATE:  15-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5525495man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-120-827-1


  Query Match             12.9%;  Score 78;  DB 1;  Length 485;
  Best Local Similarity   25.9%;  Pred. No. 0.18;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
           | |     :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      13 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 66

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      67 QQQQTQQ----AVAAAAAAVTQQLQQQQQAVVAQQAVVQQQQQQAAAV 110


RESULT   3
US-08-478-675-1
; Sequence 1, Application US/08478675
; Patent No. 5773246
;  GENERAL INFORMATION:
;    APPLICANT:  KEENE, JACK D.
;    APPLICANT:  KING, PETER H.
;    APPLICANT:  LEVINE, TODD
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS USEFUL IN THE
;    TITLE OF INVENTION:  RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
;    TITLE OF INVENTION:  INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
;    NUMBER OF SEQUENCES:  101
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 Jefferson Davis Highway, Fourth Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,675
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/120,827
;      FILING DATE:  15-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773246man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  714-158-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)413-3000
;      TELEFAX:  (703)413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  485 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-478-675-1


  Query Match             12.9%;  Score 78;  DB 1;  Length 485;
  Best Local Similarity   25.9%;  Pred. No. 0.18;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
           | |     :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      13 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 66

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      67 QQQQTQQ----AVAAAAAAVTQQLQQQQQAVVAQQAVVQQQQQQAAAV 110


RESULT   4
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT 1NFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-769-309A-5


  Query Match             11.6%;  Score 70;  DB 1;  Length 1780;
  Best Local Similarity   29.3%;  Pred. No. 11;
  Matches   17;  Conservative   10;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLE 112
           |  |  :|| : :: |:  ||: ||  : | |         |:|  |  :  | :| |
Db     224 AASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEEKQEKEPSKSAE 281


RESULT   5
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-994-570-5


  Query Match             11.6%;  Score 70;  DB 3;  Length 1780;
  Best Local Similarity   29.3%;  Pred. No. 11;
  Matches   17;  Conservative   10;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLE 112
           |  |  :|| : :: |:  ||: ||  : | |         |:|  |  :  | :| |
Db     224 AASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEEKQEKEPSKSAE 281


RESULT   6
US-08-041-538-2
; Sequence 2, Application US/08041538
; Patent No. 5385831
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R
;    APPLICANT:  Hagen, Frederick S
;    APPLICANT:  Houamed, Khaled M
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN COUPLED GLUTAMATE
;    TITLE OF INVENTION:  RECEPTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/041,538
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/648,481
;      FILING DATE:
;      APPLICATION NUMBER:  US 07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  206-623-6793
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1199 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-041-538-2


  Query Match             11.5%;  Score 69.5;  DB 1;  Length 1199;
  Best Local Similarity   32.4%;  Pred. No. 7.3;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy       9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
             :| ||  |  :  ||    :| : |      |   |:  :|| ||  |  ||    | |
Db     938 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 992

Qy      65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
                |: |   |   :| |   |:  |||::|    | |  |||
```

```
Db    993 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1033


RESULT   7
US-08-463-642-2
; Sequence 2, Application US/08463642
; Patent No. 5721107
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R
;    APPLICANT:  Hagen, Frederick S
;    APPLICANT:  Houamed, Khaled M
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN COUPLED GLUTAMATE
;    TITLE OF INVENTION:  RECEPTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,642
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;      APPLICATION NUMBER:  07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6-1-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  206-623-6793
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1199 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-463-642-2


  Query Match             11.5%;  Score 69.5;  DB 1;  Length 1199;
  Best Local Similarity   32.4%;  Pred. No. 7.3;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy      9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
          :| ||  |  :  ||    :| : |      |   |:  :|| ||  |  ||    | |
Db    938 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 992

Qy     65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| |   |:  |||::|    | |  |||
Db    993 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1033


RESULT   8
US-08-455-602-2
; Sequence 2, Application US/08455602
; Patent No. 5747267
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R
;    APPLICANT:  Hagen, Frederick S
;    APPLICANT:  Houamed, Khaled M
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN COUPLED GLUTAMATE
;    TITLE OF INVENTION:  RECEPTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  Steuart Street Tower, One Market Plaza
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,602
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;      APPLICATION NUMBER:  07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6-1-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-467-9600
;      TELEFAX:  206-623-6793
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1199 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-455-602-2


  Query Match             11.5%;  Score 69.5;  DB 1;  Length 1199;
  Best Local Similarity   32.4%;  Pred. No. 7.3;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy      9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
          :| ||  |  :  ||    :| : |      |   |:  :|| ||  |  ||    | |
Db    938 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 992

Qy     65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| |   |:  |||::|    | |  |||
Db    993 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1033


RESULT   9
US-08-465-157-2
; Sequence 2, Application US/08465157
; Patent No. 5869609
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R
;    APPLICANT:  Hagen, Frederick S
;    APPLICANT:  Houamed, Khaled M
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN COUPLED GLUTAMATE
;    TITLE OF INVENTION:  RECEPTORS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  Steuart Street Tower, One Market Plaza
```

```
;       CITY:  San Francisco
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94105
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/465,157
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/041,538
;       FILING DATE:
;       APPLICATION NUMBER:  US/07/648,481
;       FILING DATE:
;       APPLICATION NUMBER:  US 07/626,806
;       FILING DATE:  12-DEC-1990
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parmelee, Steven W
;       REGISTRATION NUMBER:  31,990
;       REFERENCE/DOCKET NUMBER:  13952-6-1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-467-9600
;       TELEFAX:  206-623-6793
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1199 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-465-157-2


  Query Match             11.5%;  Score 69.5;  DB 2;  Length 1199;
  Best Local Similarity   32.4%;  Pred. No. 7.3;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy       9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
           :| ||  |  :  ||    :| : |       |   |:  :||  ||  |  ||    | |
Db     938 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 992

Qy      65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| |   |:  |||::|    | |  |||
Db     993 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1033


RESULT  10
PCT-US91-09422-2
; Sequence 2, Application PC/TUS9109422
;  GENERAL INFORMATION:
;    APPLICANT:  Mulvihill, Eileen R.
;    APPLICANT:  Hagen, Frederick S.
;    APPLICANT:  Houamed, Khaled M.
;    APPLICANT:  Almers, Wolfhard
;    TITLE OF INVENTION:  G PROTEIN-COUPLED GLUTAMATE RECEPTORS
;    NUMBER OF SEQUENCES:  33
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/09422
;      FILING DATE:  19911212
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/672,007
;      FILING DATE:  18-MAR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/648,481
;      FILING DATE:  30-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/626,806
;      FILING DATE:  12-DEC-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parmelee, Steven W.
;      REGISTRATION NUMBER:  31,990
;      REFERENCE/DOCKET NUMBER:  13952-6PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 467-9600
;      TELEFAX:  (415) 543-5043
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1199 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US91-09422-2


  Query Match             11.5%;  Score 69.5;  DB 4;  Length 1199;
  Best Local Similarity   32.4%;  Pred. No. 7.3;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy       9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
           :| ||  |  :  ||    :| : |       |   |:  :||  ||  |  ||    | |
Db     938 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 992

Qy      65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| |   |:  |||::|    | |  |||
Db     993 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1033


RESULT  11
US-08-687-289A-6
; Sequence 6, Application US/08687289A
; Patent No. 5981195
;  GENERAL INFORMATION:
;    APPLICANT:  Fuller, Forrest H.
;    APPLICANT:  Krapcho, Karen J.
;    APPLICANT:  Hammerland, Lance G.
;    TITLE OF INVENTION:  CHIMERIC RECEPTORS AND METHODS FOR
;    TITLE OF INVENTION:  IDENTIFYING COMPOUNDS ACTIVE AT
;    TITLE OF INVENTION:  METABOTROPIC GLUTAMATE RECEPTORS AND
;    TITLE OF INVENTION:  THE USE OF SUCH COMPOUNDS IN THE
;    TITLE OF INVENTION:  TREATMENT OF NEUROLOGICAL DISORDERS
;    TITLE OF INVENTION:  AND DISEASES
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSEQ for Windows 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,289A
;      FILING DATE:  July 25, 1996
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/001,526
;      FILING DATE:  July 26, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  220/004
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1219 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Protein
US-08-687-289A-6


  Query Match             11.5%;  Score 69.5;  DB 2;  Length 1219;
  Best Local Similarity   32.4%;  Pred. No. 7.5;
  Matches   34;  Conservative   11;  Mismatches   47;  Indels   13;  Gaps    6;

Qy       9 RGDGKKVLMEARTSTSATSDVSDFD----VVFEAVSNALLVVHYHRVVPYAPVKREQPKP 64
           :| ||  |  :  ||    :| : |      |   |:  :|| || |  ||    | |
Db     958 QGSGKS-LTFSDASTKTLYNVEEEDNTPSAHFSPPSSPSMVV--HRRGP--PVATTPPLP 1012

Qy      65 AVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ 109
             |: |   |   :| | |   |:  |||::|    | |  |||
Db    1013 PHLTAEETPLFLAD--SVIPK--GLPPPLPQQQPQQPPPQQPPQQ 1053


RESULT  12
US-08-651-818A-2
; Sequence 2, Application US/08651818A
; Patent No. 5948889
;  GENERAL INFORMATION:
;    APPLICANT:  de Boer, Piet A.J.
;    APPLICANT:  Hale, Cynthia A.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR SCREENING
;    TITLE OF INVENTION:  ANTIMICROBIALS
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/651,818A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carroll, Peter G.
;      REGISTRATION NUMBER:  32,837
;      REFERENCE/DOCKET NUMBER:  CASE-02249
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  328 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-08-651-818A-2


  Query Match             11.2%;  Score 68;  DB 2;  Length 328;
  Best Local Similarity   22.8%;  Pred. No. 1.8;
  Matches   34;  Conservative   11;  Mismatches   50;  Indels   54;  Gaps    4;

Qy       6 SALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRV---------- 51
           |  |    | :   |   |   || | : | |        |  |||
Db      34 SMFRDRPLKRMKSKRDDDSYDEDVEDDEGVGE--------VRVHRVNHAPANAQEHEAAR 85

Qy      52 --------VPYAPVKREQP----------------------KPAVKQDEQKPKRQASHWA 81
                   |||  :  ||                      :|| :   ::| :|
Db      86 PSPQHQYQPPYASAQPRQPVQQPPEAQVPPQHAPHPAQPVQQPAYQPQPEQPLQQPVSPQ 145

Qy      82 VKPTAVGVHVPLPKKQEALEPAQSV--PQ 108
           | |    ||      |:| :||: |  ||
Db     146 VAPAPQPVHSAPQPAQQAFQPAEPVAAPQ 174


RESULT  13
US-08-080-255-7
; Sequence 7, Application US/08080255
; Patent No. 5487970
;  GENERAL INFORMATION:
;    APPLICANT:  Rowley, Janet D.
;    APPLICANT:  Diaz, Manuel O.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/080,255
;      FILING DATE:  19930617
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-080-255-7


  Query Match             11.1%;  Score 67;  DB 1;  Length 1400;
  Best Local Similarity   23.5%;  Pred. No. 18;
  Matches   23;  Conservative   14;  Mismatches   49;  Indels   12;  Gaps    3;

Qy      18 EARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQA 77
           | :: ||   |  :  ||   | ::       :  | |   || | |     |  |::
Db     440 EKKSKTSEKKDSKESSVVKNVVDSS------QKPTPSA---REDPAPKKSSSEPPPRKPV 490
```

```
Qy      78 SHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKA 115
           |     |  | |: ::|  ||     : : : |
Db     491 EE---KSEEGNVSAPGPESKQATTPASRKSSKQVSQPA 525


RESULT  14
US-08-465-713-7
; Sequence 7, Application US/08465713
; Patent No. 6121419
;  GENERAL INFORMATION:
;    APPLICANT:  Rowley, Janet D.
;    APPLICANT:  Diaz, Manuel O.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;    TITLE OF INVENTION:  DETECTING GENE REARRANGEMENTS AND
;    TITLE OF INVENTION:  TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/465,713
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/080,255
;      FILING DATE:  17 JUNE 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-465-713-7


  Query Match             11.1%;  Score 67;  DB 3;  Length 1400;
  Best Local Similarity   23.5%;  Pred. No. 18;
  Matches   23;  Conservative   14;  Mismatches   49;  Indels   12;  Gaps    3;

Qy      18 EARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQA 77
           | :: ||   |  :  ||   | ::       :  | |   || | |     |  |::
Db     440 EKKSKTSEKKDSKESSVVKNVVDSS------QKPTPSA---REDPAPKKSSSEPPPRKPV 490

Qy      78 SHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKA 115
           |     |  | |: ::|  ||     : : : |
Db     491 EE---KSEEGNVSAPGPESKQATTPASRKSSKQVSQPA 525


RESULT  15
PCT-US93-05857-7
; Sequence 7, Application PC/TUS9305857
;  GENERAL INFORMATION:
;    APPLICANT:  Board of Regents
;    APPLICANT:  The University of Texas System
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR DETECTING
;    TITLE OF INVENTION:  GENE REARRANGEMENTS AND TRANSLOCATIONS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P. O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05857
;      FILING DATE:  19930617
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/900,689
;      FILING DATE:  17/06/92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  ARCD:072/PAR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1400 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
PCT-US93-05857-7


  Query Match             11.1%;  Score 67;  DB 4;  Length 1400;
  Best Local Similarity   23.5%;  Pred. No. 18;
  Matches   23;  Conservative   14;  Mismatches   49;  Indels   12;  Gaps    3;

Qy      18 EARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQA 77
           | :: ||   |  :  ||   | ::       :  | |   || | |     |  |::
Db     440 EKKSKTSEKKDSKESSVVKNVVDSS------QKPTPSA---REDPAPKKSSSEPPPRKPV 490

Qy      78 SHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKA 115
           |     |  | |: ::|  ||     : : : |
Db     491 EE---KSEEGNVSAPGPESKQATTPASRKSSKQVSQPA 525


Search completed: January 31, 2001, 15:26:27
Job time: 36 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:51 ; Search time 60.61 Seconds
(without alignments)
134.435 Million cell updates/sec

Title: US-09-301-906-5
Perfect score: 605
Sequence: 1 VSGSVSALRGDGKKVLMEAR.....EPAQSVPQQSLEEKAALTFG 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83.5	13.8	282	2	A75473	probable sigma fac
2	80.5	13.3	3010	1	A45573	genome polyprotein
3	79	13.1	1545	2	T26589	hypothetical prote
4	78	12.9	483	2	A33130	elav protein - fru
5	77.5	12.8	330	2	E83084	rod shape-determin
6	76.5	12.6	733	2	A45301	microtubule-associ
7	76.5	12.6	860	2	T23296	hypothetical prote
8	74.5	12.3	2359	2	T03094	A-kinase anchor pr
9	73.5	12.1	209	2	I50223	enhancer binding p
10	73.5	12.1	1020	1	QFHUH	neurofilament trip
11	73.5	12.1	1272	2	T30248	fragile X mental r
12	73	12.1	1110	2	I51116	NF-180 - sea lampr
13	72.5	12.0	307	2	S30432	hypothetical prote
14	72.5	12.0	438	2	S77655	hypothetical prote
15	72.5	12.0	592	1	LLBY	actin-binding prot
16	72.5	12.0	691	2	A25704	synapsin I - rat
17	72.5	12.0	1072	2	A37127	microtubule-associ
18	72	11.9	630	2	S55645	transcription cont
19	71.5	11.8	417	2	T49847	hypothetical prote
20	71.5	11.8	465	2	S46759	hypothetical prote
21	71.5	11.8	511	2	T04814	dihydrolipoamide S
22	71.5	11.8	906	2	JC5963	stable tubule only
23	71.5	11.8	1102	2	JC6316	probable protein k
24	71.5	11.8	3011	1	GNWVCH	genome polyprotein
25	71	11.7	210	2	A25550	histone H1 - sea u
26	71	11.7	285	2	E64687	siderophore-mediat
27	71	11.7	491	2	A32584	Abd-B protein - fr
28	71	11.7	791	2	C82940	hypothetical prote
29	71	11.7	1474	2	T20488	hypothetical prote
30	71	11.7	2649	2	T51023	hypothetical prote
31	70.5	11.7	455	2	T38275	hypothetical prote
32	70.5	11.7	1011	2	T29806	hypothetical prote
33	70	11.6	537	2	A46611	myosin-binding pro
34	70	11.6	606	2	A43427	neurofilament trip
35	70	11.6	699	2	I38073	nucleolar phosphop
36	70	11.6	1317	2	B83346	probable non-ribos
37	70	11.6	1684	2	JW0057	gravin - human
38	69.5	11.5	261	2	JN0747	histone H1-I - Vol
39	69.5	11.5	270	2	H83619	hypothetical prote
40	69.5	11.5	617	2	T08565	hypothetical prote
41	69.5	11.5	1199	2	A41939	G protein-coupled
42	69.5	11.5	1462	2	T06819	DNA topoisomerase
43	69.5	11.5	3011	1	S40770	genome polyprotein
44	69.5	11.5	7962	2	I38346	elastic titin - hu
45	69	11.4	497	1	CEECFY	cell division prot

ALIGNMENTS

RESULT 1
A75473
probable sigma factor - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: A75473
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: A75473
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-282 <WHI>
A;Cross-references: GB:AE001935; GB:AE000513; NID:g6458517; PIDN:AAF10382.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0804
A;Map position: 1

Query Match 13.8%; Score 83.5; DB 2; Length 282;
Best Local Similarity 26.6%; Pred. No. 0.76;
Matches 33; Conservative 28; Mismatches 40; Indels 23; Gaps 7;

```
Qy       4 SVSALRGDGKKVLMEARTSTSA-----TSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVK 58
           ::| | |  :: | :|     |     |:: :: | :|  :|:         :   |||
Db     171 AISRLPGLAQEALAKAAVPAPAPPAEPTAEAAE-DDLFPTLSS---------LAATAPVT 220

Qy      59 REQP-KPAVKQDEQKPKRQASHWAVKPTA----VGVHVPLPKKQ-EALEPAQSVPQQSLE 112
           | :| :|  : ::::|: :|   || | |    : |  |||  : || :||:  |: :
Db     221 RVEPWEPPRRPEQREPETEAE--AVLPEASEATLPVSEPLPNAEPEATQPAEQSPETTAP 278

Qy     113 EKAA 116
           |  |
Db     279 EVQA 282
```

RESULT 2
A45573
genome polyprotein - hepatitis C virus (strain JT)
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstru
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C;Species: hepatitis C virus
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 16-Jun-2000
C;Accession: A45573
R;Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikat
Virus Res. 23, 39-53, 1992
A;Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier
A;Reference number: A45573; MUID:92295714

A;Accession: A45573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3010 <TAN>
A;Cross-references: GB:D11168; GB:D01171; NID:g221612; PIDN:BAA01943.1; PID:g221613
A;Experimental source: HCV-JT
A;Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C;Superfamily: hepatitis C virus genome polyprotein; DEAD/H box helicase homology
C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transm
F;2-115/Product: capsid protein C #status predicted <CPC>
F;116-191/Product: envelope protein M #status predicted <EPM>
F;192-389/Product: major envelope protein E #status predicted <MEE>
F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F;1007-1615/Product: hepacivirin #status predicted <NS3>
F;1230-1493/Domain: DEAD/H box helicase homology <DEAD>
F;1230-1237/Region: nucleotide-binding motif A (P-loop)
F;1312-1317/Region: nucleotide-binding motif B
F;1316-1319/Region: DEXH motif
F;1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F;1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 13.3%; Score 80.5; DB 1; Length 3010;
Best Local Similarity 21.1%; Pred. No. 22;
Matches 26; Conservative 24; Mismatches 64; Indels 9; Gaps 3;

```
Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       :  |  |:  |
Db    2237 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KKFPPALPIWAR 2290

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | :  :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    2291 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 2350

Qy     118 TFG 120
           |||
Db    2351 TFG 2353
```

RESULT 3
T26589
hypothetical protein Y32B12B.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26589
R;Gardner, A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z20239
A;Accession: T26589
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1545 <WIL>
A;Cross-references: EMBL:AL031632; PIDN:CAA21009.1; GSPDB:GN00023; CESP:Y32B12B.4
A;Experimental source: clone Y32B12B
C;Genetics:
A;Gene: CESP:Y32B12B.4
A;Map position: 5
A;Introns: 49/2; 311/2; 879/2; 1116/2; 1257/3; 1404/2; 1545/2

Query Match 13.1%; Score 79; DB 2; Length 1545;
Best Local Similarity 26.6%; Pred. No. 14;
Matches 29; Conservative 15; Mismatches 41; Indels 24; Gaps 4;

```
Qy      11 DGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDE 70
           | | |  : |:|:|  |:      | :|  :|          | |  || : :| |   :
Db    1323 DAKTVAPKKRSSSSMRSE--SLPAVIQAAPDA----------PEAHPKRIKTEPLVAPPQ 1370

Qy      71 QKPKRQASHWAVKPTAVGVHVPLPKKQ---EALEPAQSVPQQSLEEKAA 116
           :||          : |  | |  | :     |||   :| : : | ||
Db    1371 EKP---------STSEVRDHAPCLKMEIVDRVLEPPADLPSEKIPETAA 1410
```

RESULT 4
A33130
elav protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 24-Sep-1999
C;Accession: A33130; B33130; A30030
R;Robinow, S.; Campos, A.R.; Yao, K.M.; White, K.
Science 242, 1570-1572, 1988
A;Title: The elav gene product of Drosophila, required in neurons, has three RNP cons
A;Reference number: A33130; MUID:89072740
A;Accession: A33130
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-483 <RO1>
A;Cross-references: GB:M21153
A;Experimental source: strain Canton-S
A;Accession: B33130
A;Molecule type: mRNA
A;Residues: 1-483 <RO2>
A;Cross-references: GB:M21152; NID:g157339; PIDN:AAA28506.1; PID:g157340
C;Comment: This protein, an RNA binding protein expressed in all neurons, is required
C;Genetics:
A;Gene: FlyBase:elav
A;Cross-references: FlyBase:FBgn0000570
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleopro
F;150-230/Domain: ribonucleoprotein repeat homology <RRM1>
F;249-317/Domain: ribonucleoprotein repeat homology <RRM2>
F;403-470/Domain: ribonucleoprotein repeat homology <RRM3>

Query Match 12.9%; Score 78; DB 2; Length 483;
Best Local Similarity 25.9%; Pred. No. 4.8;
Matches 28; Conservative 22; Mismatches 48; Indels 10; Gaps 2;

```
Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
           | |      :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      11 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 64

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      65 QQQQTQQ----AVAAAAAAVTQQLQQQQQQAVVAQQAVVQQQQQQAAAV 108
```

RESULT 5
E83084
rod shape-determining protein MreC PA4480 [imported] - Pseudomonas aeruginosa (strain
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C;Accession: E83084
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950
A;Accession: E83084
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <STO>
A;Cross-references: GB:AE004862; GB:AE004091; NID:g9950716; PIDN:AAG07868.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: mreC; PA4480

Query Match 12.8%; Score 77.5; DB 2; Length 330;
Best Local Similarity 24.1%; Pred. No. 3.5;
Matches 32; Conservative 21; Mismatches 53; Indels 27; Gaps 5;

```
Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEA----------VSNALLVVH-------YHRV 51
           . | |    :| |   : |:|: : |::  :          |:    |:|         |
```

```
Db    180 GTGNPERLELR-YVADTADIKEGDLLVSSGLGQRFPAGYPVATVKEVIHDSGQPFAVVRA 238

Qy     52 VPYAPVKREQPKPAVKQDEQKPKR------QASHWAVKPTAVGVHVPLPKKQEALEPAQS 105
          || | : | :    |  | : |::      :|   | |  | |   | |  | |  |: :
Db    239 VPTAKMNRSRYVLLVFSDSRTPEQRANDAAEAQEEADKKAAAGAQAPQPAAQPAAAPSPA 298

Qy    106 VP--QQSLEEKAA 116
           |  | : :: ||
Db    299 TPAAQGAAQQPAA 311


RESULT   6
A45301
microtubule-associated protein tau - mouse
N;Alternate names: microtubule binding protein tau
C;Species: Mus musculus (house mouse)
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 13-Aug-1999
C;Accession: A45301; S31658
R;Couchie, D.; Mavilia, C.; Georgieff, I.S.; Liem, R.K.; Shelanski, M.L.; Nunez, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 4378-4381, 1992
A;Title: Primary structure of high molecular weight tau present in the peripheral nervou
A;Reference number: A45301; MUID:92262443
A;Accession: A45301
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-733 <COU>
A;Note: this sequence is inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:102045, NCBIP:102046)
R;Kenner, L.; Forstner, M.; Hutter, H.; Hoefler, G.; Kurzbauer, R.; Zatloukal, K.; Krisp
submitted to the EMBL Data Library, May 1992
A;Description: First observation of mRNA for a tau-protein from murine liver and kidney.
A;Reference number: S31658
A;Accession: S31658
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 'T',529-651 <KEN>
A;Cross-references: EMBL:Z12133; NID:g54263; PIDN:CAA78121.1; PID:g388534
C;Superfamily: microtubule-associated protein tau; MAP2/tau repeat homology
C;Keywords: microtubule binding; tandem repeat
F;544-574/Domain: MAP2/tau repeat homology <MT1>
F;575-605/Domain: MAP2/tau repeat homology <MT2>
F;606-636/Domain: MAP2/tau repeat homology <MT3>
F;637-668/Domain: MAP2/tau repeat homology <MT4>


  Query Match             12.6%;  Score 76.5;  DB 2;  Length 733;
  Best Local Similarity   28.8%;  Pred. No. 11;
  Matches   19;  Conservative   10;  Mismatches   32;  Indels    5;  Gaps    1;

Qy     57 VKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPL-----PKKQEALEPAQSVPQQSL 111
          :|   |:|      ::|  : |     |||  |  ||     | || | :|   :|:
Db     32 LKESPPQPPADDGAEEPGSETSDAKSTPTAEDVTAPLVDERAPDKQAAAQPHTEIPEGIT 91

Qy    112 EEKAAL 117
           |:| :
Db     92 AEEAGI 97


RESULT   7
T23296
hypothetical protein T09E8.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T23296; T24752
R;Percy, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19722
A;Accession: T23296
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-860 <WIL>
A;Cross-references: EMBL:Z81096; PIDN:CAB54265.1; GSPDB:GN00023; CESP:T09E8.1a
A;Experimental source: clone K03H4
R;Steward, C.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19931
A;Accession: T24752
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-860 <WI2>
A;Cross-references: EMBL:Z78065; PIDN:CAB54298.1; GSPDB:GN00023; CESP:T09E8.1a
A;Experimental source: clone T09E8
C;Genetics:
A;Gene: CESP:T09E8.1a
A;Map position: 5
A;Introns: 76/1; 124/3; 195/1; 252/2; 306/2; 394/3; 446/1; 486/1; 700/3; 817/3


  Query Match             12.6%;  Score 76.5;  DB 2;  Length 860;
  Best Local Similarity   28.6%;  Pred. No. 13;
  Matches   30;  Conservative   13;  Mismatches   33;  Indels   29;  Gaps    4;

Qy     13 KKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQK 72
          ||||     |:|:|||:|                  || :|   |      |  | :|:|
Db     32 KKVL--TSNSSSSTSDIS----------------MHRSIPEIYV------PNDKYNEEK 66

Qy     73 PKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQ----QSLEE 113
          | |: ||  ::    |         :   ||  :||:    ||| :
Db     67 PIRRKSHGNIRLRKRGDTSRRDAVEAGFEPRDTVPRCHSTQSLRD 111


RESULT   8
T03094
A-kinase anchor protein DAKAP550 - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03094
R;Han, J.D.; Baker, N.E.; Rubin, C.S.
J. Biol. Chem. 272, 26611-26619, 1997
A;Title: Molecular characterization of a novel A kinase anchor protein from drosophil
A;Reference number: Z14835; MUID:97476266
A;Accession: T03094
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2359 <HAN>
A;Cross-references: EMBL:AF003622; NID:g2393879; PIDN:AAB83959.1; PID:g2393880
A;Experimental source: strain Canton S
C;Genetics:
A;Map position: X


  Query Match             12.3%;  Score 74.5;  DB 2;  Length 2359;
  Best Local Similarity   20.7%;  Pred. No. 64;
  Matches   28;  Conservative   25;  Mismatches   47;  Indels   35;  Gaps    3;

Qy     18 EARTS-------TSATSDVSDFDVVFEAVSNALLVVHYHRV------------------ 51
          | :||       |: |   :  ::| | |  ||::|
Db   1390 EKQTSEEPEDVETAETLAAAAKEIVQEVVEAALVMVQEESTQEKPEKGANSEEEKNEIGK 1449

Qy     52 ------VPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLP---KKQEALEP 102
                :   |   |  :  :: | :||:    | :|:|    :  | |  ::|:: |
Db   1450 EEILLQLEEKPASTEVQETKIEGDLKKPEDPKGHSSVEPKTPNLEEPKPQETEQQKSQEV 1509

Qy    103 AQSVPQQSLEEKAAL 117
          |: :||:  |:  |:
Db   1510 AEELPQKPEEQVVAI 1524


RESULT   9
I50223
enhancer binding protein - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I50223
R;Ruddell, A.; Bowers, W.J.
J. Virol. 66, 6578-6586, 1992
```

```
A;Title: a1/EBP: A leucine zipper protein that binds CCAAT/enhancer elements in the avia
A;Reference number: I50223; MUID:93021386
A;Accession: I50223
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-209 <RUD>
A;Cross-references: GB:M95573; NID:g211731; PIDN:AAA48756.1; PID:g211732
C;Genetics:
A;Gene: a1/EBP


  Query Match             12.1%;  Score 73.5;  DB 2;  Length 209;
  Best Local Similarity   24.6%;  Pred. No. 5.1;
  Matches   34;  Conservative   21;  Mismatches   40;  Indels   43;  Gaps    6;

Qy       9 RGDGKKVLMEA---RTSTSATSDVSDFDVVFEAVSNALLVVHYH----------RVVPYA 55
           || |:::   |   |  ||  :  :|        :| : |:|            ::|| :
Db      46 RGAGERLRPHAFSHRRKTSPQNTATD--------ANGVSVIHTQAHSSGLQQVPQLVPVS 97

Qy      56 PVKREQPKPAVKQ----------DEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQ 104
           |    : |  ||           || : :|: :: ||| :       |  ||:| :  |
Db      98 PGGCGKAVPPSKQGKKNSFVDRNSDEYRQRRERNNMAVKKSR------LKSKQKAQDTLQ 151

Qy     105 SVPQ-----QSLEEKAAL 117
             | |     : || |  |
Db     152 RVTQLKEENERLEAKIKL 169


RESULT  10
QFHUH
neurofilament triplet H protein - human
N;Alternate names: neurofilament protein, 112K
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
C;Accession: S00979
R;Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
EMBO J. 7, 1947-1955, 1988
A;Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
A;Reference number: S00979; MUID:88328981
A;Accession: S00979
A;Molecule type: DNA
A;Residues: 1-1020 <LEE>
A;Cross-references: EMBL:X15306; NID:g35028; PIDN:CAA33366.1; PID:g1841430
A;Note: it is uncertain whether Met-1 or Met-2 is the initiator
C;Genetics:
A;Gene: GDB:NEFH
A;Cross-references: GDB:120225; OMIM:162230
A;Map position: 22q12.1-22q13.1
A;Introns: 295/1; 361/3; 403/2
C;Superfamily: neurofilament triplet H protein
C;Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
F;1-100/Domain: amino-terminal <NTD>
F;101-410/Domain: rod #status predicted <ROD>
F;411-1020/Domain: carboxyl-terminal <CTD>
F;502-826/Region: 14-residue repeats
F;503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,64
 (covalent) #status predicted
F;732,768/Binding site: phosphate (Thr) (covalent) #status predicted


  Query Match             12.1%;  Score 73.5;  DB 1;  Length 1020;
  Best Local Similarity   31.4%;  Pred. No. 31;
  Matches   22;  Conservative   10;  Mismatches   27;  Indels   11;  Gaps    2;

Qy      55 APVKREQ-------PKPAVKQDEQK----PKRQASHWAVKPTAVGVHVPLPKKQEALEPA 103
           :||| |:        :|  | :|:|    || :    : |  |      | || :|  |||
Db     822 SPVKEEEKPQEVKVKEPPKKAEEEKAPATPKTEEKKDSKKEEAPKKEAPKPKVEEKKEPA 881

Qy     104 QSVPQQSLEE 113
             |::|  |
Db     882 VEKPKESKVE 891


RESULT  11
T30248
fragile X mental retardation protein 2 - mouse
N;Alternate names: fmr2 protein
C;Species: Mus musculus (house mouse)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30248
R;Chakrabarti, L.; Bristulf, J.; Foss, G.S.; Davies, K.E.
Hum. Mol. Genet. 7, 441-448, 1998
A;Title: Expression of the murine homologue of FMR2 in mouse brain and during develop
A;Reference number: Z20786; MUID:98133924
A;Accession: T30248
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1272 <CHA>
A;Cross-references: EMBL:AJ001549; NID:g2832399; PIDN:CAA04821.1; PID:g2832400
A;Experimental source: brain
C;Genetics:
A;Gene: fmr2
A;Note: fmr2 expression in an embryo at 11 days is evident to the roof of the hind br


  Query Match             12.1%;  Score 73.5;  DB 2;  Length 1272;
  Best Local Similarity   31.0%;  Pred. No. 40;
  Matches   22;  Conservative    9;  Mismatches   23;  Indels   17;  Gaps    4;

Qy      52 VPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPK---KQEALEPAQSVPQ 108
           ||  | | :      |:   ||||| :       |  |  | ||   | :  | |: :|:
Db     777 VPVTPAKTD-----YKETASKPKRQTA-------ATAVEKPAPKGKRKHKPAETAEKIPE 824

Qy     109 --QSLEEKAAL 117
             | ||:   :
Db     825 KKQRLEDNTTI 835


RESULT  12
I51116
NF-180 - sea lamprey
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A;Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation
A;Reference number: I51116; MUID:95287814
A;Accession: I51116
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1110 <JAC>
A;Cross-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549
C;Superfamily: neurofilament triplet H protein


  Query Match             12.1%;  Score 73;  DB 2;  Length 1110;
  Best Local Similarity   35.4%;  Pred. No. 38;
  Matches   23;  Conservative   11;  Mismatches   21;  Indels   10;  Gaps    4;

Qy      56 PVKREQPKPAVKQDEQKPKRQASHWAVKPT-AVGVHVPLPK-----KQEALEPAQSVPQQ 109
           ||::|: :|     |::||: |   | ||  |     | ||     | |  :|||: |
Db     852 PVEKEE-EPEESPTEEEPKKPA---AAKPAKAPAKPKPAPKAEAEEKPEPAKPAQAKPAP 907

Qy     110 SLEEK 114
           : ||:
Db     908 AAEEE 912


RESULT  13
S30432
hypothetical protein - Streptomyces clavuligerus plasmid pSCL
C;Species: Streptomyces clavuligerus
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
C;Accession: S30432
R;Wu, X.; Roy, K.L.
```

J. Bacteriol. 175, 37-52, 1993
A;Title: Complete nucleotide sequence of a linear plasmid from Streptomyces clavuligerus
A;Reference number: S30400; MUID:93106972
A;Accession: S30432
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-307 <WUX>
A;Cross-references: EMBL:X54107
C;Genetics:
A;Genome: plasmid pSCL

Query Match 12.0%; Score 72.5; DB 2; Length 307;
Best Local Similarity 34.2%; Pred. No. 9.8;
Matches 25; Conservative 8; Mismatches 33; Indels 7; Gaps 2;

```
Qy      51 VVPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEP------AQ 104
           | | |  :||    |      ||  :| |: || :  | || |  |: :  |         :|
Db     181 VGPDAGGEREAAAPGAGVAEQSAERIAARWAERRAAAGV-VTSPGRVAAFAREAGWLLSQ 239

Qy     105 SVPQQSLEEKAAL 117
           :| | ||  ||:
Db     240 GLPAQRLEAAAAV 252
```

RESULT 14
S77655
hypothetical protein o438 - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 11-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 22-Oct-1999
C;Accession: S77655; S49517
R;Fsihi, H.; Cole, S.T.
Mol. Microbiol. 16, 909-919, 1995
A;Title: The Mycobacterium leprae genome: systematic sequence analysis identifies key ca
A;Reference number: S77652; MUID:96059637
A;Accession: S77655
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-438 <FSI>
A;Cross-references: EMBL:Z46257; NID:g559905; PIDN:CAA86359.1; PID:g559908
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1994
C;Genetics:
A;Start codon: GTG

Query Match 12.0%; Score 72.5; DB 2; Length 438;
Best Local Similarity 25.5%; Pred. No. 15;
Matches 27; Conservative 15; Mismatches 39; Indels 25; Gaps 5;

```
Qy      16 LMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRV---VPYAPVKREQPKPAVKQDEQ- 71
           |:   | | ||   :               | ||:    ||     : | || :
Db     321 LLPVETHTRATPHPA---------------HPHRLRLTTPYLHPASDTPYPACRHQRSA 364

Qy      72 -KPKRQASHWAVKPTAVGVHVPLPKKQEA-LEPAQSVPQQSLEEKA 115
            ||   ||  : ||  |    ||::::  | | ::| :::  | |
Db     365 LAPKPAQSHLTLPPTRSG---SLPRQEQTRLPPGRTVNEETRPEVA 407
```

RESULT 15
LLBY
actin-binding protein ABP1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YCR088w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1991 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: S19503; S19767; S07608
R;Dusterhoft, A.; Erdmann, D.; Hegemann, J.; Philippsen, P.; Schweitzer, B.; Spiegelberg
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19351
A;Accession: S19503
A;Molecule type: DNA
A;Residues: 1-476 <DUS>
A;Cross-references: EMBL:X59720; GSPDB:GN00003; MIPS:YCR088w
R;Frontali, L.; Grisanti, P.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19504
A;Accession: S19767
A;Molecule type: DNA
A;Residues: 14-592 <FRO>
A;Cross-references: EMBL:X59720; GSPDB:GN00003; MIPS:YCR088w
R;Drubin, D.G.; Mulholland, J.; Zhu, Z.; Botstein, D.
Nature 343, 288-290, 1990
A;Title: Homology of a yeast actin-binding protein to signal transduction proteins an
A;Reference number: S07608; MUID:90136906
A;Accession: S07608
A;Molecule type: DNA
A;Residues: 1-57,'S',59-311,'I',313-592 <DRU>
A;Cross-references: EMBL:X51780; NID:g3321; PIDN:CAA36075.1; PID:g3322
C;Genetics:
A;Gene: SGD:ABP1; MIPS:YCR088w
A;Cross-references: SGD:S0000684; MIPS:YCR088w
A;Map position: 3R
C;Superfamily: actin-binding protein ABP1; SH3 homology
C;Keywords: actin binding; duplication
F;158-180,378-401/Region: duplication
F;539-588/Domain: SH3 homology <SH3>

Query Match 12.0%; Score 72.5; DB 1; Length 592;
Best Local Similarity 31.1%; Pred. No. 21;
Matches 19; Conservative 14; Mismatches 11; Indels 17; Gaps 3;

```
Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEK 114
           |||::|:|    :|:|  |   : :           :| ||::|| | |   |:: :||:
Db     460 APVQKEEP----EQEEIAPSLPSRN----------SIPAPKQEEAPEQA---PEEEIEEE 502

Qy     115 A 115
           |
Db     503 A 503
```

Search completed: January 31, 2001, 15:27:32
Job time: 101 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:27:45 ; Search time 30.65 Seconds
(without alignments)
126.437 Million cell updates/sec

Title: US-09-301-906-5
Perfect score: 605
Sequence: 1 VSGSVSALRGDGKKVLMEAR.....EPAQSVPQQSLEEKAALTFG 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	80.5	13.3	3010	1	POLG_HCVJT	Q00269	h genome po
2	78	12.9	483	1	ELAV_DROME	P16914	drosophila
3	74	12.2	1311	1	FMR2_HUMAN	P51816	homo sapien
4	73.5	12.1	1020	1	NFH_HUMAN	P12036	homo sapien
5	72.5	12.0	592	1	ABP1_YEAST	P15891	saccharomyc
6	72.5	12.0	827	1	MOF_DROME	O02193	drosophila
7	72.5	12.0	1072	1	MAP4_BOVIN	P36225	bos taurus
8	71.5	11.8	465	1	YHU6_YEAST	P38845	saccharomyc
9	71.5	11.8	3011	1	POLG_HCVH	P27958	h genome po
10	71	11.7	205	1	H1E_STRPU	P19375	strongyloce
11	71	11.7	210	1	H1_LYTPI	P06144	lytechinus
12	71	11.7	285	1	TONB_HELPY	O25899	helicobacte
13	70	11.6	232	1	H1B_CHITE	P40276	chironomus
14	70	11.6	537	1	MYPH_CHICK	Q05623	gallus gall
15	70	11.6	1242	1	KPB1_RAT	Q64649	rattus norv
16	70	11.6	1781	1	AKAC_HUMAN	Q02952	homo sapien
17	69.5	11.5	260	1	H11_VOLCA	Q08864	volvox cart
18	69.5	11.5	1199	1	MGR1_RAT	P23385	rattus norv
19	69.5	11.5	1462	1	TOP2_PEA	O24308	pisum sativ
20	69	11.4	497	1	FTSY_ECOLI	P10121	escherichia
21	68.5	11.3	188	1	CUP8_DROME	P27780	drosophila
22	68.5	11.3	419	1	FAAA_RAT	P25093	rattus norv
23	68.5	11.3	432	1	TAU_RAT	P19332	rattus norv
24	68.5	11.3	599	1	HMW3_MYCGE	Q57081	mycoplasma
25	68.5	11.3	845	1	NFM_RAT	P12839	rattus norv
26	68.5	11.3	1527	1	MRP3_HUMAN	O15438	homo sapien
27	68	11.2	328	1	ZIPA_ECOLI	P77173	escherichia
28	68	11.2	441	1	TAU_HUMAN	P10636	homo sapien
29	68	11.2	1008	1	GALY_KLULA	P32257	kluyveromyc
30	67.5	11.2	213	1	H13_RABIT	P02251	oryctolagus
31	67.5	11.2	244	1	TONB_VIBCH	O52042	vibrio chol
32	67.5	11.2	704	1	NP14_RAT	P41777	rattus norv
33	67.5	11.2	882	1	IF2_BORBU	O51741	borrelia bu
34	67.5	11.2	1612	1	TP2B_CRILO	Q64399	cricetulus
35	67	11.1	270	1	TONB_HAEIN	P42872	haemophilus
36	67	11.1	407	1	Z174_HUMAN	Q15697	homo sapien
37	67	11.1	842	1	IF2_TREPA	O83861	treponema p
38	67	11.1	915	1	NFM_HUMAN	P07197	homo sapien
39	67	11.1	2468	1	MAPB_HUMAN	P46821	homo sapien
40	67	11.1	3969	1	HRX_HUMAN	Q03164	homo sapien
41	66.5	11.0	741	1	IF2_BACST	P04766	bacillus st
42	66.5	11.0	784	1	IF2_ENTFC	P18311	enterococcu
43	66.5	11.0	1626	1	TP2B_HUMAN	Q02880	homo sapien
44	66.5	11.0	3011	1	POLG_HCV1	P26664	h genome po
45	66	10.9	187	1	PABA_ECOLI	P00903	escherichia

ALIGNMENTS

RESULT 1
POLG_HCVJT
ID POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; D11168; BAA01943.1; -.
DR PIR; A45573; A45573.
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR INTERPRO; IPR000745; -.
DR INTERPRO; IPR001490; -.
DR INTERPRO; IPR002166; -.
DR INTERPRO; IPR002518; -.
DR INTERPRO; IPR002519; -.
DR INTERPRO; IPR002521; -.
DR INTERPRO; IPR002522; -.
DR INTERPRO; IPR002531; -.

```
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF01560; HCV_NS1; 1.
DR   PFAM; PF01538; HCV_NS2; 1.
DR   PFAM; PF01006; HCV_NS4a; 1.
DR   PFAM; PF01001; HCV_NS4b; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01543; HCV_capsid; 1.
DR   PFAM; PF01542; HCV_core; 1.
DR   PFAM; PF01539; HCV_env; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    115       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN       116    191       MATRIX PROTEIN (POTENTIAL).
FT   CHAIN       192    383       MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT   CHAIN       384    729       NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT   CHAIN       730   1006       NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT   CHAIN      1007   1615       PROTEASE/HELICASE NS3 (POTENTIAL).
FT   CHAIN      1616   1862       NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT   CHAIN      1863   2013       NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT   CHAIN      2014   3010       RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2041   2041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2077   2077       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2240   2240       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2529   2529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2788   2788       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3010 AA;  326573 MW;  94A1C77435D642BB CRC64;
```

```
  Query Match             13.3%;  Score 80.5;  DB 1;  Length 3010;
  Best Local Similarity   21.1%;  Pred. No. 8.4;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       :  |  |:  |
Db    2237 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KKFPPALPIWAR 2290

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
                | :  :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    2291 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 2350

Qy     118 TFG 120
           |||
Db    2351 TFG 2353
```

```
RESULT   2
ELAV_DROME
ID   ELAV_DROME     STANDARD;      PRT;   483 AA.
AC   P16914;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ELAV PROTEIN (EMBRYONIC LETHAL ABNORMAL VISUAL PROTEIN).
GN   ELAV.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CANTON-S;
RX   MEDLINE=89072740; PubMed=3144044;
RA   Robinow S., Campos A.R., Yao K.-M., White K.;
RT   "The elav gene product of Drosophila, required in neurons, has three
RT   RNP consensus motifs.";
RL   Science 242:1570-1572(1988).
RN   [2]
RP   ERRATUM.
RA   Robinow S., Campos A.R., Yao K.-M., White K.;
RL   Science 243:12-12(1989).
CC   -!- FUNCTION: IS REQUIRED FOR THE PROPER DEVELOPMENT AND MAINTENANCE
CC       OF NEURONS PRESUMABLY BY AFFECTING RNA METABOLISM.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).
CC   -!- SIMILARITY: BELONGS TO THE ELAV FAMILY OF RNP PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M21152; AAA28506.1; -.
DR   PIR; A33130; A33130.
DR   HSSP; P19339; 2SXL.
DR   FLYBASE; FBgn0000570; elav.
DR   INTERPRO; IPR000504; -.
DR   INTERPRO; IPR002343; -.
DR   PFAM; PF00076; rrm; 3.
DR   PRINTS; PR00961; HUDSXLRNA.
DR   PROSITE; PS50102; RRM; 3.
DR   PROSITE; PS00030; RRM_RNP_1; 2.
KW   RNA-binding; Repeat.
FT   DOMAIN       24    126       ALA/GLN-RICH.
FT   DOMAIN      203    210       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      290    297       RNA-BINDING (RNP1) (BY SIMILARITY).
FT   DOMAIN      443    450       RNA-BINDING (RNP1) (BY SIMILARITY).
SQ   SEQUENCE   483 AA;  50817 MW;  E5733610631 0810B CRC64;
```

```
  Query Match             12.9%;  Score 78;  DB 1;  Length 483;
  Best Local Similarity   25.9%;  Pred. No. 1.8;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
             | |      :   | :| : |:  :     | ||        |   | ::::| : |: |
Db      11 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 64

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      65 QQQQTQQ----AVAAAAAAVTQQLQQQQQAVVAQQAVVQQQQQQAAAV 108
```

```
RESULT   3
FMR2_HUMAN
ID   FMR2_HUMAN     STANDARD;      PRT;  1311 AA.
AC   P51816; Q13521; Q14323; P78407;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
```

```
DE   FRAGILE X MENTAL RETARDATION 2 PROTEIN (PROTEIN FMR-2) (FMR2P) (OX19
DE   PROTEIN) (FRAGILE X E MENTAL RETARDATION SYNDROME PROTEIN).
GN   FMR2 OR OX19.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FETAL BRAIN, AND PLACENTA;
RX   MEDLINE=96241583; PubMed=8673085;
RA   Gecz J., Gedeon A.K., Sutherland G.R., Mulley J.C.;
RT   "Identification of the gene FMR2, associated with FRAXE mental
RT   retardation.";
RL   Nat. Genet. 13:105-108(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=96241584; PubMed=8673086;
RA   Gu Y., Shen Y., Gibbs R.A., Nelson D.L.;
RT   "Identification of FMR2, a novel gene associated with the FRAXE CCG
RT   repeat and CpG island.";
RL   Nat. Genet. 13:109-113(1996).
RN   [3]
RP   SEQUENCE FROM N.A. (POTENTIAL SHORT ISOFORM).
RC   TISSUE=FETAL BRAIN;
RX   MEDLINE=96422267; PubMed=8824884;
RA   Chakrabarti L., Knight S.J.L., Flannery A.V., Davies K.E.;
RT   "A candidate gene for mild mental handicap at the FRAXE fragile
RT   site.";
RL   Hum. Mol. Genet. 5:275-282(1996).
RN   [4]
RP   SEQUENCE OF 520-1311 FROM N.A.
RA   Dugan-Rocha S.P., Gibbs R.A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS: IT IS POSSIBLE THAT A NUMBER OF FORMS ARE
CC       PRODUCED BY ALTERNATIVE SPLICING OF FMR2.
CC   -!- TISSUE SPECIFICITY: BRAIN (MOST ABUNDANT IN HIPPOCAMPUS AND
CC       AMYGDALA), PLACENTA AND LUNG.
CC   -!- DISEASE: DEFECTS IN FMR2 ARE THE CAUSE OF AN X-LINKED FORM OF
CC       MENTAL RETARDATION CALLED FRAXE. LOSS OF FMR2 EXPRESSION IS
CC       CORRELATED WITH FRAXE CCG(N) EXPANSION. NORMAL INDIVIDUALS HAVE 6-
CC       35 COPIES OF THE REPEAT, WHEREAS CYTOGENETICALLY POSITIVE,
CC       DEVELOPMENTALLY DELAYED MALES HAVE MORE THAN 200 COPIES AND SHOW
CC       METHYLATION OF THE ASSOCIATED CPG ISLAND.
CC   -!- SIMILARITY: TO HUMAN LAF-4 AND AF-4 (MLLT2) PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U48436; AAC82513.1; -.
DR   EMBL; L76569; AAA99416.1; -.
DR   EMBL; X95463; CAA64730.1; -.
DR   EMBL; AC005731; AAB69976.1; -.
DR   MIM; 309548; -.
KW   Triplet repeat expansion; Disease mutation; Alternative splicing.
FT   VARSPLIC    364    392       MISSING (IN POTENTIAL ISOFORM 2).
FT   VARSPLIC    466    466       S -> K (IN POTENTIAL SHORT ISOFORM).
FT   VARSPLIC    467   1311       MISSING (IN POTENTIAL SHORT ISOFORM).
FT   CONFLICT     57     60       MISSING (IN REF. 2 AND 3).
FT   CONFLICT    195    195       D -> A (IN REF. 3).
FT   CONFLICT    416    421       MISSING (IN REF. 3).
FT   CONFLICT    470    470       A -> V (IN REF. 2).
FT   CONFLICT    548    548       P -> Q (IN REF. 4).
FT   CONFLICT    970    971       MISSING (IN REF. 2 AND 4).
SQ   SEQUENCE   1311 AA;  144739 MW;  9F151EC06BB636CB CRC64;


  Query Match             12.2%;  Score 74;  DB 1;  Length 1311;
  Best Local Similarity   32.4%;  Pred. No. 14;
  Matches   23;  Conservative   12;  Mismatches   28;  Indels    8;  Gaps    3;

Qy      50 RVVPYAPVKREQPKPAVKQDEQKPKRQASHWAV-KPTAVGVHVPLPKKQEALEPAQSVPQ 108
           ||  :: :    ||  |:  ||||| :  || ||   |      :| : :| |: :|:
Db     812 RVPGHSSLHAAPAKPDHKETATKPKRQTAVTAVEKPAPKG-----KRKHKPIEVAEKIPE 866

Qy     109 --QSLEEKAAL 117
             | |||   :
Db     867 KKQRLEEATTI 877


RESULT   4
NFH_HUMAN
ID   NFH_HUMAN      STANDARD;      PRT;  1020 AA.
AC   P12036;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)
DE   (NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H).
GN   NEFH OR NFH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88328981; PubMed=3138108;
RA   Lees J.F., Shneidman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT   "The structure and organization of the human heavy neurofilament
RT   subunit (NF-H) and the gene encoding it.";
RL   EMBO J. 7:1947-1955(1988).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X15306; CAA33366.1; -.
DR   EMBL; X15307; CAA33366.1; JOINED.
DR   EMBL; X15308; CAA33366.1; JOINED.
DR   EMBL; X15309; CAA33366.1; JOINED.
DR   PIR; S00979; QFHUH.
DR   MIM; 162230; -.
DR   INTERPRO; IPR001664; -.
DR   PFAM; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW   Phosphorylation.
FT   DOMAIN        1    100       HEAD.
FT   DOMAIN      101    413       ROD.
FT   DOMAIN      414   1020       TAIL.
FT   DOMAIN      101    132       COIL 1A.
FT   DOMAIN      133    145       LINKER 1.
FT   DOMAIN      146    244       COIL 1B.
FT   DOMAIN      245    266       LINKER 12.
```

```
FT   DOMAIN      267    288       COIL 2A.
FT   DOMAIN      289    292       LINKER 2.
FT   DOMAIN      293    413       COIL 2B.
SQ   SEQUENCE   1020 AA;  111780 MW;  1177C9DCB3DCF1D4 CRC64;


  Query Match             12.1%;  Score 73.5;  DB 1;  Length 1020;
  Best Local Similarity   31.4%;  Pred. No. 12;
  Matches   22;  Conservative   10;  Mismatches   27;  Indels   11;  Gaps    2;

Qy      55 APVKREQ-------PKPAVKQDEQK----PKRQASHWAVKPTAVGVHVPLPKKQEALEPA 103
           :||| |:        :|  | :|:|    || :    : |  |     | || :|  |||
Db     822 SPVKEEEKPQEVKVKEPPKKAEEEKAPATPKTEEKKDSKKEEAPKKEAPKPKVEEKKEPA 881

Qy     104 QSVPQQSLEE 113
           |::|  |
Db     882 VEKPKESKVE 891


RESULT   5
ABP1_YEAST
ID   ABP1_YEAST     STANDARD;      PRT;   592 AA.
AC   P15891;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ACTIN BINDING PROTEIN.
GN   ABP1 OR YCR088W OR YCR88W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE-90136906; PubMed-2405279;
RA   Drubin D.G., Mulholland J., Zhu Z., Botstein D.;
RT   "Homology of a yeast actin-binding protein to signal transduction
RT   proteins and myosin-I.";
RL   Nature 343:288-290(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Frontali L., Grisanti P.;
RL   Submitted (MAR-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY BE INVOLVED IN THE SPATIAL ORGANIZATION OF CELL
CC       SURFACE GROWTH. AN OVERPRODUCTION OF ABP1 CAUSES THE ASSEMBLY OF
CC       THE CORTICAL ACTIN SKELETON AT INAPPROPRIATE SITES ON THE CELL
CC       SURFACE, RESULTING IN DELOCALIZED SURFACE GROWTH.
CC   -!- SUBCELLULAR LOCATION: CORTICAL CYTOSKELETON.
CC   -!- SIMILARITY: TO SIGNAL TRANSDUCTION PROTEINS AND MYOSIN I.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51780; CAA36075.1; -.
DR   EMBL; X59720; CAA42253.1; -.
DR   PIR; S19503; LLBY.
DR   HSSP; P29355; 1SEM.
DR   SGD; S0000684; ABP1.
DR   INTERPRO; IPR001452; -.
DR   INTERPRO; IPR002108; -.
DR   PFAM; PF00018; SH3; 1.
DR   PFAM; PF00241; cofilin_ADF; 1.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50002; SH3; 1.
KW   Cytoskeleton; Actin-binding; SH3 domain; Repeat.
FT   SIMILAR       1    142       TO YEAST COFILIN.
FT   DOMAIN       86     97       ACTIN-BINDING (POTENTIAL).
FT   DOMAIN      532    592       SH3.
FT   DOMAIN      200    575       3 X 10 AA APPROXIMATE REPEATS.
FT   REPEAT      200    209       1.
FT   REPEAT      436    445       2.
FT   REPEAT      566    575       3.
FT   CONFLICT     58     58       L -> S (IN REF. 1).
FT   CONFLICT    312    312       K -> I (IN REF. 1).
SQ   SEQUENCE   592 AA;  65576 MW;  39523510704D94AA CRC64;


  Query Match             12.0%;  Score 72.5;  DB 1;  Length 592;
  Best Local Similarity   31.1%;  Pred. No. 7.9;
  Matches   19;  Conservative   14;  Mismatches   11;  Indels   17;  Gaps    3;

Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEK 114
           |||::|:|    :|:|  |   : :           :| ||::|| | |   |:: :|||:
Db     460 APVQKEEP----EQEEIAPSLPSRN----------SIPAPKQEEAPEQA---PEEEIEEE 502

Qy     115 A 115
           |
Db     503 A 503


RESULT   6
MOF_DROME
ID   MOF_DROME      STANDARD;      PRT;   827 AA.
AC   O02193; Q9W463;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MALES-ABSENT ON THE FIRST PROTEIN (EC 2.3.1.-) (PUTATIVE ACETYL
DE   TRANSFERASE MOF).
GN   MOF.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE-97299882; PubMed-9155031;
RA   Hilfiker A., Hilfiker-Kleiner D., Pannuti A., Lucchesi J.C.;
RT   "mof, a putative acetyl transferase gene related to the Tip60 and MOZ
RT   human genes and to the SAS genes of yeast, is required for dosage
RT   compensation in Drosophila.";
RL   EMBO J. 16:2054-2060(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE-20196006; PubMed-10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: PROBABLE HISTONE ACETYLTRANSFERASE THAT PLAYS A DIRECT
CC       ROLE IN THE SPECIFIC HISTONE ACETYLATION ASSOCIATED WITH DOSAGE
CC       COMPENSATION. DOSAGE COMPENSATION INSURES THAT MALES WITH A SINGLE
CC       X CHROMOSOME HAVE THE SAME AMOUNT OF MOST X-LINKED GENE PRODUCTS
CC       AS FEMALES WITH TWO X CHROMOSOMES. MAY BE DIRECTLY INVOLVED IN THE
CC       ACETYLATION OF HISTONE 4 AT LYS 16 ON THE X CHROMOSOME OF
CC       DROSOPHILA MALES.
CC   -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U71219; AAC47507.1; -.
DR   EMBL; AE003435; AAF46095.1; -.
DR   FLYBASE; FBgn0014340; mof.
DR   INTERPRO; IPR002717; -.
DR   PFAM; PF01853; MOZ_SAS; 1.
KW   Transferase; Acyltransferase; Zinc-finger.
FT   ZN_FING      572    594       C2HC-TYPE.
FT   SIMILAR      678    701       TO SOME ACETYLTRANSFERASES.
FT   DOMAIN       112    116       POLY-SER.
FT   DOMAIN       135    141       POLY-GLN.
FT   DOMAIN       245    252       POLY-PRO.
FT   DOMAIN       340    346       POLY-GLU.
FT   DOMAIN       466    472       POLY-ALA.
FT   VARIANT      691    691       G -> E (MALES FAIL TO METAMORPHOSE AND
FT                                 HATCH).
SQ   SEQUENCE   827 AA;  92656 MW;  8B9C5E87E3B01317 CRC64;


  Query Match             12.0%;  Score 72.5;  DB 1;  Length 827;
  Best Local Similarity   33.0%;  Pred. No. 12;
  Matches   31;  Conservative   11;  Mismatches   25;  Indels   27;  Gaps    5;

Qy      21 TSTSAT--SDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQAS 78
           ||||:|  |  | :| | ||                || : |  :|  :| :|: |::
Db     105 TSTSSTRSSSSSRYDDVSEAEE-------------APPEPEPEQP--QQQQQEEKKEDG 148

Qy      79 HWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLE 112
              ||        | | : || ||||  ||:  |
Db     149 QDQVKS-------PGPVELEAQEPAQ--PQKQKE 173


RESULT   7
MAP4_BOVIN
ID   MAP4_BOVIN     STANDARD;      PRT;  1072 AA.
AC   P36225;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   MICROTUBULE-ASSOCIATED PROTEIN 4 (MICROTUBULE-ASSOCIATED PROTEIN-U)
DE   (MAP-U).
GN   MAP4.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90338002; PubMed=2380192;
RA   Aizawa H., Emori Y., Murofushi H., Kawasaki H., Sakai H.,
RA   Suzuki K.;
RT   "Molecular cloning of a ubiquitously distributed
RT   microtubule-associated protein with Mr 190,000.";
RL   J. Biol. Chem. 265:13849-13855(1990).
RN   [2]
RP   DOMAINS.
RX   MEDLINE=91236765; PubMed=2033072;
RA   Aizawa H., Emori Y., Mori A., Murofushi H., Sakai H., Suzuki K.;
RT   "Functional analyses of the domain structure of
RT   microtubule-associated protein-4 (MAP-U).";
RL   J. Biol. Chem. 266:9841-9846(1991).
CC   -!- FUNCTION: NON-NEURONAL MICROTUBULE-ASSOCIATED PROTEIN. PROMOTES
CC       MICROTUBULE ASSEMBLY.
CC   -!- TISSUE SPECIFICITY: IS DISTRIBUTED UBIQUITOUSLY AMONG ALL TISSUES
CC       BUT AMOUNTS ARE LOWER IN CEREBELLUM AND LIVER.
CC   -!- PTM: PHOSPHORYLATION OF THE PRO-RICH REGION IN THE C-TERMINUS
CC       NEGATIVELY REGULATES MAP-4 ACTIVITY TO PROMOTE MICROTUBULE
CC       ASSEMBLY.
CC   -!- SIMILARITY: CONTAINS 3 TAU/MAP REPEATS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90149; BAA14179.1; -.
DR   PIR; A37127; A37127.
DR   INTERPRO; IPR001084; -.
DR   PFAM; PF00418; tubulin-binding; 3.
DR   PROSITE; PS00229; TAU_MAP_1; 3.
KW   Microtubules; Repeat; Phosphorylation.
FT   DOMAIN       244    530       19 X 14 AA TANDEM REPEATS.
FT   REPEAT       244    257       1.
FT   REPEAT       258    271       2.
FT   REPEAT       272    285       3.
FT   REPEAT       286    299       4.
FT   REPEAT       300    313       5.
FT   REPEAT       314    327       6.
FT   REPEAT       328    341       7.
FT   REPEAT       342    355       8.
FT   REPEAT       384    391       9 (INCOMPLETE).
FT   REPEAT       392    405       10.
FT   REPEAT       406    417       11.
FT   REPEAT       418    431       12.
FT   REPEAT       432    445       13.
FT   REPEAT       446    460       14.
FT   REPEAT       461    474       15.
FT   REPEAT       475    488       16.
FT   REPEAT       489    502       17.
FT   REPEAT       503    516       18.
FT   REPEAT       517    530       19.
FT   REPEAT       907    937       TAU/MAP MOTIF.
FT   REPEAT       938    968       TAU/MAP MOTIF.
FT   REPEAT       969   1000       TAU/MAP MOTIF.
SQ   SEQUENCE  1072 AA;  111914 MW;  E8C17A730989F0D2 CRC64;


  Query Match             12.0%;  Score 72.5;  DB 1;  Length 1072;
  Best Local Similarity   25.4%;  Pred. No. 16;
  Matches   35;  Conservative   15;  Mismatches   61;  Indels   27;  Gaps    5;

Qy       6 SALRGDGKKVLMEARTS-----------TSATSDVSDFDVVFEAVSNALLVVHYHRVVP 53
           : | |  || :  | |             ||  | :   |   : |: |   :   || |
Db     678 TTLGGSNKKPMSLASGSVPAAPPKRPAAATSRPSTLPSKDTKPKPVAEA--KIPEKRVSP 735
```

```
Qy      54 YAPVKREQPKPAVKQDEQKPKRQASHWAVKP--TAVGVHVPLPKKQEAL----EPAQ--- 104
           |      ||  |  :  ||  |:     |  |: :  ||||:  |:    :||:
Db     736 SKPASAPAVKPGSKSTQAVPKAPATATLASPGSTSRNLSTPLPKRPTAIKTEGKPAEIKK 795

Qy     105 ----SVPQQSLEEKAALT 118
               | |      |:  |
Db     796 MATKSAPADLSRPKSTTT 813


RESULT   8
YHU6_YEAST
ID   YHU6_YEAST     STANDARD;      PRT;   465 AA.
AC   P38845;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 51.1 KDA PROTEIN IN DCD1-MRPL6 INTERGENIC REGION.
GN   YHR146W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC   Saccharomycetaceae; Saccharomyces.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SIMILARITY: TO YEAST YNL173C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10397; AAB68982.1; -.
DR   PIR; S46759; S46759.
DR   SGD; S0001189; YHR146W.
KW   Hypothetical protein.
SQ   SEQUENCE   465 AA;  51115 MW;  30880758F37991C7 CRC64;


  Query Match             11.8%;  Score 71.5;  DB 1;  Length 465;
  Best Local Similarity   25.9%;  Pred. No. 7.5;
  Matches   38;  Conservative   18;  Mismatches   52;  Indels   39;  Gaps    8;

Qy       2 SGSVSALRGDGKKVLMEARTSTSATSDVS---------DFDV-VFEAVSNALLVVHYHRV 51
           |  |  :  |  :  |:| | ||    | |         | :|   : : |
Db     323 SSKVEPVEGSLQSKLVEKRESTEGVLDGSKKVENKAKKDEEVFTLDPIVNK--------- 373

Qy      52 VPYAPVKREQ-----PKPAVKQDEQKPKRQ----------ASHWAVKPTAVGVH---VPL 93
             |  |:  ||       ||| ::::| |:|           :    ||:|  |    |
Db     374 APKLPLTDEQTAEGRKSPAVSEEKEKKKKQEKGSKEVKRSETSKEKKPSAKEVKKQTVKA 433

Qy      94 PKKQEALEPAQSVPQQSLEEKAALTFG 120
           |||| |  |  |  ::  ::|    ||
Db     434 PKKQTA-SPLSSSTEEPKKKKTGF-FG 458


RESULT   9
POLG_HCVH
ID   POLG_HCVH      STANDARD;      PRT;  3011 AA.
AC   P27958;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN) (P22);
DE   ENVELOPE GLYCOPROTEIN E1 (GP32) (GP35); ENVELOPE GLYCOPROTEIN E2
DE   (GP68) (GP70) (NS1); PROTEIN P7; NONSTRUCTURAL PROTEIN NS2 (P21)
DE   (EC 3.4.22.-); PROTEASE/HELICASE NS3 (P70) (EC 3.4.21.-);
DE   NONSTRUCTURAL PROTEIN NS4A (P4); NONSTRUCTURAL PROTEIN NS4B (P27);
DE   NONSTRUCTURAL PROTEIN NS5A (P56); NONSTRUCTURAL PROTEIN NS5B (P66)
DE   (P70) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48)].
OS   Hepatitis C virus (isolate H) (HCV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92052256; PubMed=1658800;
RA   Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M.,
RA   Prince A.M.;
RT   "Genomic structure of the human prototype strain H of hepatitis C
RT   virus: comparison with American and Japanese isolates.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
RX   MEDLINE=97331322; PubMed=9187654;
RA   Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
RT   "Structure of the hepatitis C virus RNA helicase domain.";
RL   Nat. Struct. Biol. 4:463-467(1997).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
RX   MEDLINE=98154321; PubMed=9493270;
RA   Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
RA   Murcko M.A., Lin C., Caron P.R.;
RT   "Hepatitis C virus NS3 RNA helicase domain with a bound
RT   oligonucleotide: the crystal structure provides insights into the mode
RT   of unwinding.";
RL   Structure 6:89-100(1998).
CC   -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
CC   -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
CC       NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
CC   -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
CC       ACTIVATION OF NS3.
CC   -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
CC   -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
CC       ESSENTIAL ROLE IN THE VIRUS REPLICATION.
CC   -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC       LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
CC       AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
CC   -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
CC       PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
CC   -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY C18.
CC   -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M67463; AAA45534.1; -.
DR   PIR; A36814; GNWVCH.
DR   PDB; 1HEI; 25-NOV-98.
DR   PDB; 1A1V; 16-FEB-99.
DR   MEROPS; S29.001; -.
DR   MEROPS; U39.001; -.
DR   INTERPRO; IPR000745; -.
DR   INTERPRO; IPR001490; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002518; -.
DR   INTERPRO; IPR002519; -.
DR   INTERPRO; IPR002521; -.
```

```
DR   INTERPRO; IPR002522; -.
DR   INTERPRO; IPR002531; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF01560; HCV_NS1; 1.
DR   PFAM; PF01538; HCV_NS2; 1.
DR   PFAM; PF01006; HCV_NS4a; 1.
DR   PFAM; PF01001; HCV_NS4b; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01543; HCV_capsid; 1.
DR   PFAM; PF01542; HCV_core; 1.
DR   PFAM; PF01539; HCV_env; 1.
KW   Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW   Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW   Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW   3D-structure.
FT   INIT_MET      1      1       REMOVED FROM CAPSID PROTEIN C BY THE
FT                                CELLULAR AMINOPEPTIDASE.
FT   CHAIN         1    191       CAPSID PROTEIN C.
FT   CHAIN       192    383       ENVELOPE GLYCOPROTEIN E1.
FT   CHAIN       384    746       ENVELOPE GLYCOPROTEIN E2.
FT   CHAIN       747    809       PROTEIN P7.
FT   CHAIN       810   1026       NONSTRUCTURAL PROTEIN NS2.
FT   CHAIN      1027   1657       PROTEASE/HELICASE NS3.
FT   CHAIN      1658   1711       NONSTRUCTURAL PROTEIN NS4A.
FT   CHAIN      1712   1972       NONSTRUCTURAL PROTEIN NS4B.
FT   CHAIN      1973   2420       NONSTRUCTURAL PROTEIN NS5A.
FT   CHAIN      2421   3011       NONSTRUCTURAL PROTEIN NS5B.
FT   TRANSMEM    347    369       POTENTIAL.
FT   ACT_SITE   1083   1083       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1107   1107       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1165   1165       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   NP_BIND    1230   1237       ATP (POTENTIAL).
FT   SITE       1316   1319       DECH BOX.
FT   CARBOHYD    196    196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    209    209       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    305    305       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    430    430       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    476    476       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    556    556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    576    576       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    623    623       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    645    645       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3011 AA;  327142 MW;  772CBB29CCD94753 CRC64;
```

```
  Query Match             11.8%;  Score 71.5;  DB 1;  Length 3011;
  Best Local Similarity   22.0%;  Pred. No. 66;
  Matches   27;  Conservative   23;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       |  |  ||  |
Db    2237 MGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAEILRKS------RRFAPALPVWAR 2290

Qy      60 EQPKPAVKQDEQKPKRQ--ASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | :  :  :||  :    |    |      || |:|:  :   :|    :| | |
Db    2291 PDYNPLLVETWKKPDYEPPVVHGCPLPPPRSPPVPPPRKKRTVVLTESTLPTALAELATK 2350

Qy     118 TFG 120
           :||
Db    2351 SFG 2353
```

```
RESULT  10
H1E_STRPU
ID   H1E_STRPU      STANDARD;      PRT;   205 AA.
AC   P19375;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HISTONE H1, EARLY EMBRYONIC.
OS   Strongylocentrotus purpuratus (Purple sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC   Strongylocentrotus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82265580; PubMed=7107576;
RA   Levy S., Sures I., Kedes L.;
RT   "The nucleotide and amino acid coding sequence of a gene for H1
RT   histone that interacts with euchromatin. The early embryonic H1 gene
RT   of the sea urchin Strongylocentrotus purpuratus.";
RL   J. Biol. Chem. 257:9438-9443(1982).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- DEVELOPMENTAL STAGE: EARLY EMBRYONIC.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01354; CAA24644.1; -.
DR   HSSP; P02259; 1HST.
DR   INTERPRO; IPR001386; -.
DR   PFAM; PF00538; linker_histone; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   205 AA;  22173 MW;  FF553FA7B9C33853 CRC64;
```

```
  Query Match             11.7%;  Score 71;  DB 1;  Length 205;
  Best Local Similarity   34.3%;  Pred. No. 3.2;
  Matches   23;  Conservative    9;  Mismatches   31;  Indels    4;  Gaps    2;

Qy      52 VPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVP--LPKKQEALEPAQSVPQQ 109
           |  || | :  ||  |  |:| |::    | |  |      |   |||  | :|  : |::
Db     138 VKAAPKKAK--KPVKKTTEKKEKKKTPKKAPKKPAAKKSTPKKTPKKAAAKKPKTAKPKK 195

Qy     110 SLEEKAA 116
             :|||
Db     196 PAXKKAA 202
```

```
RESULT  11
H1_LYTPI
ID   H1_LYTPI       STANDARD;      PRT;   210 AA.
AC   P06144;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   LATE HISTONE H1.
OS   Lytechinus pictus (Painted sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae;
OC   Lytechinus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=87040778; PubMed=3022245;
RA   Knowles J.A., Childs G.J.;
RT   "Comparison of the late H1 histone genes of the sea urchins
RT   Lytechinus pictus and Strongelocentrotus purpuratus.";
RL   Nucleic Acids Res. 14:8121-8133(1986).
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
```

```
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X04488; CAA28177.1; -.
DR   PIR; A25550; A25550.
DR   HSSP; P02259; 1HST.
DR   INTERPRO; IPR001386; -.
DR   PFAM; PF00538; linker_histone; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   210 AA;  21746 MW;  08C38F64894007E2 CRC64;


  Query Match             11.7%;  Score 71;  DB 1;  Length 210;
  Best Local Similarity   31.7%;  Pred. No. 3.3;
  Matches   20;  Conservative   11;  Mismatches   20;  Indels   12;  Gaps    2;

Qy      56 PVKREQPKPAVKQDEQKP--KRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEE 113
           | |:   ||| |:  :||  |: |   | || |          ::| :||:   ::   :
Db     156 PAKKAAKKPAAKKAAKKPAAKKPAKKAAKKPAA----------KKAAKPAKKAAKKPAAK 205

Qy     114 KAA 116
           |||
Db     206 KAA 208


RESULT  12
TONB_HELPY
ID   TONB_HELPY     STANDARD;      PRT;   285 AA.
AC   O25899;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   TONB PROTEIN.
GN   TONB OR HP1341.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC       CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC       THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC       TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC       REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC       RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC       MEMBRANE PROTEINS (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC       MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC       PERIPLASM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000635; AAD08383.1; -.
DR   TIGR; HP1341; -.
KW   Transport; Protein transport; Inner membrane; Periplasmic;
KW   Transmembrane; Repeat.
FT   DOMAIN        1     14       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     15     35       ANCHOR (POTENTIAL).
FT   DOMAIN       36    285       PERIPLASMIC (POTENTIAL).
SQ   SEQUENCE   285 AA;  31585 MW;  3897A8BB7B70BDF8 CRC64;


  Query Match             11.7%;  Score 71;  DB 1;  Length 285;
  Best Local Similarity   25.5%;  Pred. No. 4.8;
  Matches   28;  Conservative   18;  Mismatches   48;  Indels   16;  Gaps    4;

Qy       8 LRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVK 67
             || |  : | :| | |  |  ::  :      :||           :   :|:||   |
Db      35 LREDAPEPLAQAGT-TKVTMSLASINTNSNTKTNA----------ESAKPKEEPKEKPK 82

Qy      68 QDEQK---PKRQASHWAVKPTAVGVHVPLPKKQEALEP-AQSVPQQSLEE 113
           ::| |   ||:: :    ||       | || :   ||  :  |:  :||
Db      83 KEEPKKEEPKKEVTKPKPKPKPKPKPKPKPEPKPEPKPEPKPEPKVEE 132


RESULT  13
H1B_CHITE
ID   H1B_CHITE      STANDARD;      PRT;   232 AA.
AC   P40276;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HISTONE H1B.
OS   Chironomus tentans (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Chironomoidea; Chironomidae; Chironominae; Chironomus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulze E., Wisniewski J.R., Nagel S., Gavenis K., Grossbach U.;
RL   Submitted (XXX-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC       NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29108; AAA21716.1; -.
DR   HSSP; P08287; 1GHC.
DR   INTERPRO; IPR001386; -.
DR   PFAM; PF00538; linker_histone; 1.
KW   Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ   SEQUENCE   232 AA;  24388 MW;  60E7E69694706D00 CRC64;


  Query Match             11.6%;  Score 70;  DB 1;  Length 232;
  Best Local Similarity   38.8%;  Pred. No. 4.7;
  Matches   19;  Conservative    5;  Mismatches   19;  Indels    6;  Gaps    1;

Qy      55 APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPA 103
```

```
          ||  :  |||   : | |||| |:  | || |       | : | :||
Db    186 APKPKAAPKPKAAKTETKPKRAAAPKAKKPAAE------KKPKAAKKPA 228


RESULT  14
MYPH_CHICK
ID   MYPH_CHICK     STANDARD;      PRT;   537 AA.
AC   Q05623;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   MYOSIN-BINDING PROTEIN H (MYBP-H) (H-PROTEIN) (86 KDA PROTEIN).
GN   MYBPH.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 2-37.
RC   TISSUE=PECTORALIS MUSCLE;
RX   MEDLINE=93155224; PubMed=7679114;
RA   Vaughan K.T., Weber F.E., Einheber S., Fischman D.A.;
RT   "Molecular cloning of chicken myosin-binding protein (MyBP) H (86-kDa
RT   protein) reveals extensive homology with MyBP-C (C-protein) with
RT   conserved immunoglobulin C2 and fibronectin type III motifs.";
RL   J. Biol. Chem. 268:3670-3676(1993).
CC   -!- FUNCTION: BINDS TO MYOSIN; PROBABLY INVOLVED IN INTERACTION WITH
CC       THICK MYOFILAMENTS IN THE A-BAND.
CC   -!- TISSUE SPECIFICITY: SKELETAL MUSCLE. SEEMS TO BE ALSO EXPRESSED IN
CC       THE SLOW TONIC ALD MUSCLE. NOT DETECTED IN GIZZARD OR HEART.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE MYBP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L05605; AAA21418.1; -.
DR   INTERPRO; IPR001777; -.
DR   INTERPRO; IPR002965; -.
DR   INTERPRO; IPR003006; -.
DR   PFAM; PF00041; fn3; 2.
DR   PFAM; PF00047; ig; 2.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR01217; PRICHEXTENSN.
KW   Immunoglobulin domain; Cell adhesion; Muscle protein; Thick filament;
KW   Repeat.
FT   DOMAIN      135    221       FIBRONECTIN TYPE-III.
FT   DOMAIN      253    312       IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN      331    416       FIBRONECTIN TYPE-III.
FT   DOMAIN      458    518       IG-LIKE C2-TYPE DOMAIN.
FT   CONFLICT      2      2       T -> G (IN AA SEQUENCE).
FT   CONFLICT      9      9       A -> P (IN AA SEQUENCE).
FT   CONFLICT     15     15       A -> K (IN AA SEQUENCE).
SQ   SEQUENCE   537 AA;  58678 MW;  06C4CF0EFE1DD233 CRC64;


  Query Match           11.6%;  Score 70;  DB 1;  Length 537;
  Best Local Similarity  32.3%;  Pred. No. 13;
  Matches   20;  Conservative    8;  Mismatches   28;  Indels    6;  Gaps    2;

Qy     55 APVKREQPKPAVKQ-DEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEE 113
          ||  :: | || |:  |  || : :     ||    | | ||:: |  | :      ||
Db     13 APAAKKAPAPASKKAPEPAPKEKPA-----PTPKEGHAPTPKEEHAPPPKEEHAPPPKEE 67

Qy    114 KA 115
          |
Db     68 HA 69
```

```
RESULT  15
KPB1_RAT
ID   KPB1_RAT       STANDARD;      PRT;  1242 AA.
AC   Q64649; Q64650; Q64651; Q9QZ77;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PHOSPHORYLASE B KINASE ALPHA REGULATORY CHAIN, SKELETAL MUSCLE
DE   ISOFORM (PHOSPHORYLASE KINASE ALPHA M SUBUNIT).
GN   PHKA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RA   Brushia R.J., Hoye E.R., Walsh D.A.;
RT   "Fundamental physicochemical characterization of the phosphorylase b
RT   kinase holoenzyme and alpha-gamma-delta and gamma-delta subunit
RT   subcomplexes reconstituted in baculovirus-infected insect cells.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 244-267; 547-658 AND 908-1136 FROM N.A.
RC   TISSUE=SKELETAL MUSCLE;
RX   MEDLINE=92381046; PubMed=1512265;
RA   Cawley K.C., Akita C.G., Wineinger M.A., Carlsen R.C., Gorin F.A.,
RA   Walsh D.A.;
RT   "Coordinated expression of phosphorylase kinase subunits in
RT   regenerating skeletal muscle.";
RL   J. Biol. Chem. 267:17287-17295(1992).
CC   -!- FUNCTION: PHOSPHORYLASE B KINASE CATALYZES THE PHOSPHORYLATION OF
CC       SERINE IN CERTAIN SUBSTRATES, INCLUDING TROPONIN I. THE ALPHA
CC       CHAIN MAY BIND CALMODULIN.
CC   -!- ENZYME REGULATION: BY PHOSPHORYLATION OF VARIOUS SERINE RESIDUES
CC       AND BY CALCIUM.
CC   -!- PATHWAY: GLYCOGEN METABOLISM.
CC   -!- SUBUNIT: POLYMER OF 16 CHAINS, FOUR EACH OF ALPHA, BETA, GAMMA,
CC       AND DELTA. ALPHA AND BETA ARE REGULATORY CHAINS, GAMMA IS THE
CC       CATALYTIC CHAIN, AND DELTA IS CALMODULIN.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHORYLASE B KINASE REGULATORY
CC       CHAINS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF197561; AAF06673.1; -.
DR   EMBL; M92917; AAA41858.1; -.
DR   EMBL; M92918; AAA41859.1; -.
DR   EMBL; M92919; AAA41860.1; -.
KW   Glycogen metabolism; Muscle; Phosphorylation; Calmodulin-binding;
KW   Multigene family.
FT   DOMAIN      812    842       CALMODULIN-BINDING (POTENTIAL).
FT   DOMAIN     1065   1105       CALMODULIN-BINDING (POTENTIAL).
FT   MOD_RES     974    974       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES     987    987       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1009   1009       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   MOD_RES    1020   1020       PHOSPHORYLATION (BY CAPK)
FT                                (BY SIMILARITY).
FT   MOD_RES    1022   1022       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES    1025   1025       PHOSPHORYLATION (BY SIMILARITY).
FT   CONFLICT    245    245       T -> S (IN REF. 2).
FT   CONFLICT    264    265       IS -> V (IN REF. 2).
FT   CONFLICT    926    926       I -> L (IN REF. 2).
FT   CONFLICT    957    957       R -> K (IN REF. 2).
SQ   SEQUENCE   1242 AA;  139154 MW;  15E61F40D4D2E82C CRC64;
```

Query Match 11.6%; Score 70; DB 1; Length 1242;
Best Local Similarity 25.2%; Pred. No. 33;
Matches 30; Conservative 12; Mismatches 45; Indels 32; Gaps 4;

```
Qy      21 TSTSATSDVSDFDVVFEAVSNALLVV------HYHRVVPYAPVKREQP---------KPA 65
           | ||    :  |    :   : : :|      |   |  | | |  ||         |
Db     680 TPTSRKGGLDRFRAAVQTTCDLMSLVAKAKELHIQNVHMYLPTKLFQPSRPSLNLLDSPE 739

Qy      66 VKQDFQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQ-----SLEEKAALTF 119
             || | |           :| | | ||: |      ||:  |     ||:|:| : :
Db     740 SPQDSQVP-----------SVRVEVHLPRDQSGEVDFQSLVSQLKETSSLQEQADILY 786
```

Search completed: January 31, 2001, 15:33:10
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:25:55 ; Search time 102.04 Seconds
(without alignments)
137.838 Million cell updates/sec

Title: US-09-301-906-5
Perfect score: 605
Sequence: 1 VSGSVSALRGDGKKVLMEAR.....EPAQSVPQQSLEEKAALTFG 120

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	88	14.5	462	10	Q9ZRQ1	Q9zrq1 arabidopsis
2	83.5	13.8	282	2	Q9RW65	Q9rw65 deinococcus
3	82.5	13.6	1820	4	Q9P2P6	Q9p2p6 homo sapien
4	80.5	13.3	453	12	Q9YKI4	Q9yki4 hepatitis c
5	80.5	13.3	453	12	Q9YKI3	Q9yki3 hepatitis c
6	80.5	13.3	2284	12	Q81817	Q81817 hepatitis c
7	80.5	13.3	3010	12	Q81541	Q81541 hepatitis c
8	80.5	13.3	3010	12	P89966	P89966 hepatitis c
9	79	13.1	1545	5	Q9XX31	Q9xx31 caenorhabdi
10	78.5	13.0	447	12	O89168	O89168 hepatitis c
11	78	12.9	479	5	Q9V3F6	Q9v3f6 drosophila
12	77.5	12.8	302	5	Q9VQH5	Q9vqh5 drosophila
13	77	12.7	942	13	O57452	O57452 fundulus he
14	77	12.7	1545	5	Q9VYN1	Q9vyn1 drosophila
15	76.5	12.6	385	2	O07110	O07110 enterococcu
16	76.5	12.6	430	11	Q60684	Q60684 mus musculu
17	76.5	12.6	860	5	P90906	P90906 caenorhabdi
18	76	12.6	16215	5	Q9NFS3	Q9nfs3 drosophila
19	75.5	12.5	318	10	Q9SL96	Q9sl96 arabidopsis
20	75.5	12.5	451	2	Q9JWQ1	Q9jwq1 bacillus ha
21	75	12.4	1852	5	Q9VH42	Q9vh42 drosophila
22	75	12.4	3008	12	O39929	O39929 hepatitis c
23	74.5	12.3	2359	5	O16024	O16024 drosophila
24	74.5	12.3	3614	5	Q9W4E2	Q9w4e2 drosophila
25	74	12.2	467	13	O42465	O42465 salmo salar
26	74	12.2	1272	4	O43786	O43786 homo sapien
27	74	12.2	1311	4	O60215	O60215 homo sapien
28	74	12.2	1557	5	Q9V513	Q9v513 drosophila
29	74	12.2	1718	5	Q9NE62	Q9ne62 leishmania
30	73.5	12.1	209	13	Q02612	Q02612 gallus gall
31	73.5	12.1	933	4	Q9UQ14	Q9uq14 homo sapien
32	73.5	12.1	1026	4	Q9UJS7	Q9ujs7 homo sapien
33	73.5	12.1	1179	11	Q9JK71	Q9jk71 rattus norv
34	73.5	12.1	1272	11	O55112	O55112 mus musculu
35	73.5	12.1	3010	12	Q9J3G4	Q9j3g4 hepatitis c
36	73	12.1	713	10	Q9SK77	Q9sk77 arabidopsis
37	73	12.1	1110	13	Q91255	Q91255 petromyzon
38	72.5	12.0	438	2	Q50175	Q50175 mycobacteri
39	72.5	12.0	952	2	Q9RN18	Q9rn18 bacillus an
40	72.5	12.0	3011	12	O36609	O36609 hepatitis c
41	72	11.9	630	12	Q66652	Q66652 equine herp
42	71.5	11.8	417	3	Q9P513	Q9p513 neurospora
43	71.5	11.8	511	10	Q9SZ31	Q9sz31 arabidopsis
44	71.5	11.8	906	11	O55129	O55129 mus musculu
45	71.5	11.8	956	10	Q9LJ64	Q9lj64 arabidopsis

ALIGNMENTS

RESULT 1
Q9ZRQ1
ID Q9ZRQ1 PRELIMINARY; PRT; 462 AA.
AC Q9ZRQ1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE 2-OXOGLUTARATE DEHYDROGENASE E2 SUBUNIT (EC 2.3.1.61).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. C24;
RA Machuy N., Klein M., Mueller-Roeber B.;
RT "Cloning and characterization of 2-Oxoglutarate dehydrogenase from
RT Arabidopsis thaliana.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ223803; CAA11553.1; -.
DR HSSP; P07016; 1E2O.
DR INTERPRO; IPR000089; -.
DR INTERPRO; IPR001078; -.
DR INTERPRO; IPR003016; -.
DR PFAM; PF00198; 2-oxoacid_dh; 1.
DR PFAM; PF00364; biotin_lipoyl; 1.
DR PROSITE; PS00189; LIPOYL; 1.
DR PRODOM; PD001115; -; 1.
KW Transferase; Acyltransferase; Lipoyl.
SQ SEQUENCE 462 AA; 49939 MW; 6A811153FB34EEAA CRC64;

Query Match 14.5%; Score 88; DB 10; Length 462;
Best Local Similarity 32.1%; Pred. No. 0.32;
Matches 35; Conservative 12; Mismatches 30; Indels 32; Gaps 6;

```
Qy      12 GKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD-- 69
           | ||   || ||||          :|||:            || ::   ||| |
Db     161 GNKV---ARISTSA----------DAVSHV-----------APSEKAPEKPAPKPSPP 194

Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALT 118
```

```
          :||| :::  | || |     | | || | ||   :| :  | :  :|
Db    195 AEKPKVESTKVAEKPKAPS--PPPPSKQSAKEP--QLPPKDRERRVPMT 239


RESULT   2
Q9RW65
ID   Q9RW65      PRELIMINARY;      PRT;   282 AA.
AC   Q9RW65;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   SIGMA FACTOR, PUTATIVE.
GN   DR0804.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001935; AAF10382.1; -.
DR   TIGR; DR0804; -.
SQ   SEQUENCE   282 AA;  29916 MW;  343E6915E1993DC5 CRC64;


  Query Match             13.8%;  Score 83.5;  DB 2;  Length 282;
  Best Local Similarity   26.6%;  Pred. No. 0.53;
  Matches   33;  Conservative   28;  Mismatches   40;  Indels   23;  Gaps    7;

Qy       4 SVSALRGDGKKVLMEARTSTSA-----TSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVK 58
           ::| | |  :: | :|     |     |:: :: | :|  :|:         :   |||
Db     171 AISRLPGLAQEALAKAAVPAPAPPAEPTAEAAE-DDLFPTLSS---------LAATAPVT 220

Qy      59 REQP-KPAVKQDEQKPKRQASHWAVKPTA----VGVHVPLPKKQ-EALEPAQSVPQQSLE 112
           | :| :|  : ::::|: :|   || | |    : |  |||  : || :||:  |: :
Db     221 RVEPWEPPRRPEQREPETEAE--AVLPEASEATLPVSEPLPNAEPEATQPAEQSPETTAP 278

Qy     113 EKAA 116
           |  |
Db     279 EVQA 282


RESULT   3
Q9P2P6
ID   Q9P2P6      PRELIMINARY;      PRT;  1820 AA.
AC   Q9P2P6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   KIAA1300 PROTEIN (FRAGMENT).
GN   KIAA1300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes.XVI.
RT   The complete sequences of 150 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:65-73(2000).
DR   EMBL; AB037721; BAA92538.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   1820 AA;  197962 MW;  B5AD7B415B106898 CRC64;


  Query Match             13.6%;  Score 82.5;  DB 4;  Length 1820;
  Best Local Similarity   26.3%;  Pred. No. 6;
  Matches   30;  Conservative   13;  Mismatches   38;  Indels   33;  Gaps    4;

Qy      12 GKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVK---- 67
           |  |::|   | ::|    : |                  ||  | |||  : | |
Db     876 GANVILEGLGSDTSTVSQEEGD-----------------VPGVPQKREAEETAQKMAQL 917

Qy      68 ---QDE------QKPKRQASHWAVKPTAVGVHVP--LPKKQEALEPAQSVPQQS 110
              |:|      | |   :||   :   || |:|   |   :|  |  | |::|
Db     918 LYLQEESTPYKPQSPSIPSSHLRFQKAPVGQHLPSVSPSVSDAFLPPSSQPEES 971


RESULT   4
Q9YKI4
ID   Q9YKI4      PRELIMINARY;      PRT;   453 AA.
AC   Q9YKI4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   NON-STRUCTURAL PROTEIN 5A (FRAGMENT).
GN   NS5A.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1A;
RA   Paterson M., Laxton C.D., Thomas H.C., Ackrill A.M., Foster G.R.;
RT   "The inhibition of the anti-viral effects of IFN-alpha by the NS5A
RT   protein of hepatitis C virus in human cells.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ133098; CAB37932.1; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
FT   NON_TER     453     453
SQ   SEQUENCE   453 AA;  49625 MW;  F5E3F14BA8EAA3FD CRC64;


  Query Match             13.3%;  Score 80.5;  DB 12;  Length 453;
  Best Local Similarity   22.0%;  Pred. No. 1.9;
  Matches   27;  Conservative   25;  Mismatches   62;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  |      ::  | |  |  |:  |
Db     266 MGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAE------ILRKYRRFAPALPIWAR 319

Qy      60 EQPKPAVKQDEQKPKRQ--ASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             :  | : :  :||  :    |    |      || |:|:  :   :|    :| | |
Db     320 PEYNPPLLETWKKPDYEPPVVHGCPLPPPQSPPVPPPRKKRMVVLTESTVSTALAELATK 379

Qy     118 TFG 120
           :||
Db     380 SFG 382


RESULT   5
Q9YKI3
ID   Q9YKI3      PRELIMINARY;      PRT;   453 AA.
AC   Q9YKI3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   NON-STRUCTURAL PROTEIN 5A (FRAGMENT).
GN   NS5A.
```

```
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1A;
RA   Paterson M., Laxton C.D., Thomas H.C., Ackrill A.M., Foster G.R.;
RT   "The inhibition of the anti-viral effects of IFN-alpha by the NS5A
RT   protein of hepatitis C virus in human cells.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ133099; CAB37933.1; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
FT   NON_TER     453    453
SQ   SEQUENCE   453 AA;  49614 MW;  C69D6EBEA4300BAF CRC64;


  Query Match             13.3%;  Score 80.5;  DB 12;  Length 453;
  Best Local Similarity   22.0%;  Pred. No. 1.9;
  Matches   27;  Conservative   25;  Mismatches   62;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  |      ::  | |  |  |:  |
Db     266 MGGNITRVESENKVVILDSFDPLVAEEDEREVSVPAE------ILRKYRRFAPALPIWAR 319

Qy      60 EQPKPAVKQDEQKPKRQ--ASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :  | : :  :||  :    |    |      || |:|:  :   :|    :| | |
Db     320 PEYNPPLLETWKKPDYEPPVVHGCPLPPPQSPPVPPPRKKRMVVLTESTVSTALAELATK 379

Qy     118 TFG 120
           :||
Db     380 SFG 382


RESULT   6
Q81817
ID   Q81817      PRELIMINARY;      PRT;  2284 AA.
AC   Q81817;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   POLYPROTEIN PRECURSOR.
OS   Hepatitis C virus type 2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=40271;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94068484; PubMed=7504283;
RA   Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,
RA   Akagi T., Kimura K., Shimotohno K.;
RT   "Proteolytic processing and membrane association of putative
RT   nonstructural proteins of hepatitis C virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94333810; PubMed=8056334;
RA   Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;
RT   "Identification of the domain required for trans-cleavage activity of
RT   hepatitis C viral serine proteinase.";
RL   Gene 145:215-219(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95056078; PubMed=7966638;
RA   Tanji Y., Hizikata M., Hirowatari Y., Shimotohono K.;
RT   "Hepatitis C virus polyprotein processing: kinetics and mutagenic
RT   analysis of serine proteinase-dependent cleavage.";
RL   J. Virol. 68:8418-8422(1994).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95156583; PubMed=7853491;
RA   Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.;
RT   "Hepatitis C virus-encoded nonstructural protein NS4A has versatile
RT   functions in viral protein processing.";
RL   J. Virol. 69:1575-1581(1995).
DR   EMBL; D16435; BAA03905.1; -.
DR   INTERPRO; IPR000745; -.
DR   INTERPRO; IPR001410; -.
DR   INTERPRO; IPR001490; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002518; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01001; HCV_NS4b; 1.
DR   PFAM; PF01006; HCV_NS4a; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
DR   PFAM; PF01538; HCV_NS2; 1.
KW   Signal; Polyprotein; Nonstructural protein.
FT   SIGNAL       4     20
FT   SIGNAL      55     83
FT   CHAIN       21     54
FT   CHAIN       84    300       NS2.
FT   CHAIN      301    931       NS3.
FT   CHAIN      932    985       NS4A.
FT   CHAIN      986   1246       NS4B.
FT   CHAIN     1247   1693       NS5A.
FT   CHAIN     1694   2284       NS5B.
SQ   SEQUENCE   2284 AA;  247213 MW;  DC272A1517046337 CRC64;


  Query Match             13.3%;  Score 80.5;  DB 12;  Length 2284;
  Best Local Similarity   21.1%;  Pred. No. 13;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :       :  |  |:  |
Db    1511 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KKFPPALPIWAR 1564

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | : :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    1565 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 1624

Qy     118 TFG 120
           |||
Db    1625 TFG 1627


RESULT   7
Q81541
ID   Q81541      PRELIMINARY;      PRT;  3010 AA.
AC   Q81541;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   POLYPROTEIN.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JT';
RX   MEDLINE=92295714; PubMed=1318627;
RA   Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T.,
RA   Hijikata M., Ishimura Y., Shimotohno K.;
RT   "Molecular cloning of hepatitis C virus genome from a single Japanese
RT   carrier: sequence variation within the same individual and among
RT   infected individuals.";
RL   Virus Res. 23:39-53(1992).
DR   EMBL; D11355; BAA18894.1; -.
DR   HSSP; P26663; 1JXP.
DR   INTERPRO; IPR000745; -.
DR   INTERPRO; IPR001410; -.
```

```
DR   INTERPRO; IPR001490; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002518; -.
DR   INTERPRO; IPR002519; -.
DR   INTERPRO; IPR002521; -.
DR   INTERPRO; IPR002522; -.
DR   INTERPRO; IPR002531; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01001; HCV_NS4b; 1.
DR   PFAM; PF01006; HCV_NS4a; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
DR   PFAM; PF01538; HCV_NS2; 1.
DR   PFAM; PF01539; HCV_env; 1.
DR   PFAM; PF01542; HCV_core; 1.
DR   PFAM; PF01543; HCV_capsid; 1.
DR   PFAM; PF01560; HCV_NS1; 1.
DR   PRODOM; PD186062; -; 1.
KW   Polyprotein; Envelope protein.
FT   CHAIN         1    191       PUT. P22.
FT   CHAIN       192    383       PUT. GP35.
FT   CHAIN       384    729       PUT. GP70.
FT   CHAIN       730   1006       PUT. NS2.
FT   CHAIN      1007   1615       PUT. NS3.
FT   CHAIN      1616   1862       PUT. NS4A.
FT   CHAIN      1863   2013       PUT. NS4B.
FT   CHAIN      2014   3010       PUT. NS5.
SQ   SEQUENCE   3010 AA;  326565 MW;  05F0B2102CF9DD9D CRC64;


  Query Match             13.3%;  Score 80.5;  DB 12;  Length 3010;
  Best Local Similarity   21.1%;  Pred. No. 17;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :        :  |  |:  |
Db    2237 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KKFPPALPIWAR 2290

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | :  :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    2291 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 2350

Qy     118 TFG 120
           |||
Db    2351 TFG 2353


RESULT   8
P89966
ID   P89966      PRELIMINARY;      PRT;  3010 AA.
AC   P89966;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   POLYPROTEIN.
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-TYPE 1B;
RA   Tanaka T.;
RT   "TMORF.";
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; D89872; BAA14035.1; -.
DR   HSSP; P26663; 1JXP.
DR   INTERPRO; IPR000745; -.
DR   INTERPRO; IPR001410; -.
DR   INTERPRO; IPR001490; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002518; -.
DR   INTERPRO; IPR002519; -.
DR   INTERPRO; IPR002521; -.
DR   INTERPRO; IPR002522; -.
DR   INTERPRO; IPR002531; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01001; HCV_NS4b; 1.
DR   PFAM; PF01006; HCV_NS4a; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
DR   PFAM; PF01538; HCV_NS2; 1.
DR   PFAM; PF01539; HCV_env; 1.
DR   PFAM; PF01542; HCV_core; 1.
DR   PFAM; PF01543; HCV_capsid; 1.
DR   PFAM; PF01560; HCV_NS1; 1.
DR   PRODOM; PD186062; -; 1.
KW   Polyprotein.
SQ   SEQUENCE   3010 AA;  327024 MW;  E075BD9CFD8D1261 CRC64;


  Query Match             13.3%;  Score 80.5;  DB 12;  Length 3010;
  Best Local Similarity   21.1%;  Pred. No. 17;
  Matches   26;  Conservative   24;  Mismatches   64;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :        :  |  |:  |
Db    2237 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KKFPPALPIWAR 2290

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
             | :  :  : |     | |    |   |  :| |:|:  :   :|    :| | |
Db    2291 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 2350

Qy     118 TFG 120
           |||
Db    2351 TFG 2353


RESULT   9
Q9XX31
ID   Q9XX31      PRELIMINARY;      PRT;  1545 AA.
AC   Q9XX31;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   Y32B12B.4 PROTEIN.
GN   Y32B12B.4.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Gardner A.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; AL031632; CAA21009.1; -.
SQ   SEQUENCE   1545 AA;  173298 MW;  B2898E8F623F071E CRC64;
```

```
  Query Match             13.1%;  Score 79;  DB 5;  Length 1545;
  Best Local Similarity   26.6%;  Pred. No. 12;
  Matches   29;  Conservative   15;  Mismatches   41;  Indels   24;  Gaps    4;

Qy      11 DGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDE 70
           | | |  : |:|:|  |:      | :|  :|          | |  || : :| |   :
Db    1323 DAKTVAPKKRSSSSMRSE--SLPAVIQAAPDA----------PEAHPKRIKTEPLVAPPQ 1370

Qy      71 QKPKRQASHWAVKPTAVGVHVPLPKKQ---EALEPAQSVPQQSLEEKAA 116
           :||          : |  | |  | :     |||   :| : : | ||
Db    1371 EKP---------STSEVRDHAPCLKMEIVDRVLEPPADLPSEKIPETAA 1410


RESULT  10
O89168
ID   O89168      PRELIMINARY;      PRT;   447 AA.
AC   O89168;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   POLYPROTEIN (FRAGMENT).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HCV-1BJK;
RA   Song J., Wang F., Itoh M., Hotta H.;
RT   "Partial abrogation of antiviral activity of interferon by the NS5A
RT   protein of hepatitis C virus.";
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB010249; BAA31545.1; -.
DR   INTERPRO; IPR002166; -.
DR   INTERPRO; IPR002868; -.
DR   PFAM; PF00998; HCV_RdRP; 1.
DR   PFAM; PF01506; HCV_NS5a; 1.
KW   Polyprotein.
FT   NON_TER       1      1
FT   NON_TER     447    447
SQ   SEQUENCE   447 AA;  48355 MW;  1AC0ED49CB35004B CRC64;


  Query Match             13.0%;  Score 78.5;  DB 12;  Length 447;
  Best Local Similarity   21.1%;  Pred. No. 3;
  Matches   26;  Conservative   23;  Mismatches   65;  Indels    9;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPV-KR 59
           : |::: :  : | |::::     |  |  :  |  | :  :          |  |:  |
Db     265 MGGNITRVESENKVVILDSFDPLRAEEDEREVSVAAEILRKS------KNFPPALPIWAR 318

Qy      60 EQPKPAVKQDEQKPK--RQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
              | :  :  : |     | |     |   |  :| |:|:  :   :|    :| | |
Db     319 PDYNPPLLESWKSPDYVPPAVHGCPLPPTTGPPIPPPRKKRTVVLTESTVSSALAELATK 378

Qy     118 TFG 120
           |||
Db     379 TFG 381


RESULT  11
Q9V3F6
ID   Q9V3F6      PRELIMINARY;      PRT;   479 AA.
AC   Q9V3F6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ELAV PROTEIN.
GN   ELAV OR EG:65F1.2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Madueno E., de Pablos B., Modolell J.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003417; AAF45517.1; -.
DR   EMBL; AL022139; CAA18091.1; -.
DR   HSSP; P19339; 2SXL.
DR   FLYBASE; FBgn0000570; elav.
DR   INTERPRO; IPR000504; -.
DR   INTERPRO; IPR002343; -.
DR   PFAM; PF00076; rrm; 3.
DR   PRINTS; PR00961; HUDSXLRNA.
DR   PROSITE; PS00030; RNP_1; 2.
SQ   SEQUENCE   479 AA;  50311 MW;  D5D1554716982662 CRC64;


  Query Match             12.9%;  Score 78;  DB 5;  Length 479;
  Best Local Similarity   25.9%;  Pred. No. 3.7;
  Matches   28;  Conservative   22;  Mismatches   48;  Indels   10;  Gaps    2;

Qy      10 GDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQD 69
           | |      :   | :| : |:  :     | ||        |   | ::::| : |: |
Db       7 GGGVDTQAQLMQSAAAAAAVAATNAAAAPVQNAA------AVAAAAQLQQQQVQQAILQV 60
```

```
Qy      70 EQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAAL 117
           :|:  :|    ||   |  |   | ::|:|:   |:| ||  :: ||:
Db      61 QQQQTQQ----AVAAAAAAVTQQLQQQQQAVVAQQAVVQQQQQQAAAV 104


RESULT 12
Q9VQH5
ID   Q9VQH5      PRELIMINARY;      PRT;   302 AA.
AC   Q9VQH5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG2973 PROTEIN.
GN   CG2973.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003581; AAF51198.1; -.
DR   FLYBASE; FBgn0031467; CG2973.
DR   INTERPRO; IPR000618; -.
DR   PFAM; PF00379; insect_cuticle; 1.
DR   PRINTS; PR00947; CUTICLE.
DR   PROSITE; PS00233; CUTICLE; 1.
SQ   SEQUENCE   302 AA;  32060 MW;  5DAB89E0B5D06168 CRC64;


  Query Match           12.8%;  Score 77.5;  DB 5;  Length 302;
  Best Local Similarity 39.4%;  Pred. No. 2.5;
  Matches   28;  Conservative    4;  Mismatches   28;  Indels   11;  Gaps    3;

Qy       1 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKRE 60
           | |  | :  || |     || |    ||  |: |   |  ||  |    ||| | |
Db     183 VRGFYSLIDPDGYK-----RTVTYTADDVHGFNAVVNRVPYALKAV----VVPVAQV--A 231

Qy      61 QPKPAVKQDEQ 71
           || | | :||:
Db     232 QPTPFVARDER 242


RESULT 13
O57452
ID   O57452      PRELIMINARY;      PRT;   942 AA.
AC   O57452;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ARYL HYDROCARBON RECEPTOR 1 ALPHA.
GN   AHR1 ALPHA.
OS   Fundulus heteroclitus (Killifish) (Mummichog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC   Cyprinodontiformes; Fundulidae; Fundulus.
OX   NCBI_TaxID=8078;
RN   [1]
RP   SEQUENCE OF 117-336 FROM N.A.
RX   MEDLINE=98054305; PubMed=9391097;
RA   Hahn M.E., Karchner S.I., Shapiro M.A., Perera S.A.;
RT   "Molecular evolution of two vertebrate aryl hydrocarbon (dioxin)
RT   receptors (AHR1 and AHR2) and the PAS family.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:13743-13748(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Karchner S.I., Powell W.H., Hahn M.E.;
RT   "Identification and functional characterization of two highly
RT   divergent aryl hydrocarbon receptors (AHR1 and AHR2) in the teleost
RT   Fundulus heteroclitus: Evidence for a novel subfamily of ligand-
RT   binding basic helix loop helix-per-arnt-sim (bHLH-PAS) factors.";
RL   J. Biol. Chem. 274:33814-33824(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Hahn M.E., Karchner S.I., Shapiro M.A., Perera S.A.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Hahn M.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF024591; AAC60334.2; -.
DR   INTERPRO; IPR000014; -.
DR   INTERPRO; IPR001092; -.
DR   INTERPRO; IPR001610; -.
DR   INTERPRO; IPR003015; -.
DR   PFAM; PF00785; PAC; 1.
DR   PFAM; PF00989; PAS; 2.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW   Receptor.
SQ   SEQUENCE   942 AA;  103682 MW;  E8D041E4E61BB790 CRC64;


  Query Match           12.7%;  Score 77;  DB 13;  Length 942;
  Best Local Similarity 22.0%;  Pred. No. 10;
  Matches   26;  Conservative   25;  Mismatches   47;  Indels   20;  Gaps    4;

Qy       2 SGSVSA-LRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVV--------HYHRVV 52
           ||: :|  ||:| :      |: :|   :|: ::: :|::  :||:          | :
Db      81 SGNGAAEARGNGDE------TTGAAEKQISEGELLLQALNGFVLVITADGTIFYSSHTIQ 134

Qy      53 PY-----APVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQS 105
             |     |  :     |  ::|:  |: |||: |      |      : :||  |
Db     135 DYLGFHQTDVIHQSAYELVHTEDQQELRRNLHWALNPPQAAASVISQDSSQEMEPDSS 192
```

```
RESULT 14
Q9VYN1
ID   Q9VYN1      PRELIMINARY;      PRT;  1545 AA.
AC   Q9VYN1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG11245 PROTEIN.
GN   CG11245.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003489; AAF48161.1; -.
DR   FLYBASE; FBgn0030388; CG11245.
SQ   SEQUENCE   1545 AA;  166082 MW;  2ABCEABED136C8B8 CRC64;

  Query Match             12.7%;  Score 77;  DB 5;  Length 1545;
  Best Local Similarity   24.5%;  Pred. No. 19;
  Matches   36;  Conservative   16;  Mismatches   31;  Indels   64;  Gaps    6;

Qy      19 ARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQAS 78
           ||     |:|:         ||::     | | ||    ::|||| |||| ||| : :
Db     138 ARPIAINTADID--------VSSS----RYRRAVPSKSQEKEQPKEEVKQKEQKEQAEEE 185

Qy      79 HWAVKP--------------------TAVGVHVPL-----------------PKKQE-- 98
             |||                       : :  : |:                 |||::
Db     186 ---VKPSPKETPCESSDASAETRPNRRSTIRRNRPVVRLSTIRRRSKDRSAESPKKEQVQ 242

Qy      99 ---------ALEPAQSVPQQSLEEKAA 116
                    | ||: |  |:  || ||
Db     243 EHQFRDLPPATEPSLSWRQKLAEELAA 269


RESULT 15
O07110
ID   O07110      PRELIMINARY;      PRT;   385 AA.
AC   O07110;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   CELL DIVISION PROTEIN.
GN   DIV1B.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A24836;
RX   MEDLINE=97431524; PubMed=9287029;
RA   Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E.,
RA   Dougherty T.J.;
RT   "Identification and characterization of cell wall-cell division gene
RT   clusters in pathogenic gram-positive cocci.";
RL   J. Bacteriol. 179:5632-5635(1997).
DR   EMBL; U94707; AAC45637.1; -.
KW   Cell division.
SQ   SEQUENCE   385 AA;  43741 MW;  1BFB17496D980D5A CRC64;

  Query Match             12.6%;  Score 76.5;  DB 2;  Length 385;
  Best Local Similarity   31.3%;  Pred. No. 4.2;
  Matches   21;  Conservative   15;  Mismatches   26;  Indels    5;  Gaps    2;

Qy      53 PYAPVKREQPKPAV----KQDEQKPKRQASHWAVKPTAVGVHVPLP-KKQEALEPAQSVP 107
           |  | | |: :| :    ||:::  |:||   | |       | :  ::|   || | :|
Db      16 PLPPKKSEESQPELTPWQKQNQEYLKKQAEEAASKGENEQAEVTITLQEQSQEEPQQHLP 75

Qy     108 QQSLEEK 114
           |:::||:
Db      76 QETVEEE 82


Search completed: January 31, 2001, 15:32:37
Job time: 402 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:28:18 ; Search time 42.21 Seconds
(without alignments)
1812.166 Million cell updates/sec

Title: US-09-301-906-15
Perfect score: 11520
Sequence: 1 MDYIRPLRVFSFPHVNNTLE.....QSVSDALLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_36:*
1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11520	100.0	2237	21	Y58148	GLRaV-3 polyprotei
2	5247	45.5	1390	18	W21636	Grapevine leafroll
3	1495	13.0	291	21	Y58145	GLRaV-3 polyprotei
4	1432	12.4	272	21	Y58144	GLRaV-3 polyprotei
5	794.5	6.9	2639	20	W73476	Grapevine leafroll
6	631.5	5.5	309	21	Y91969	PMWaV-2 helicase.
7	605	5.3	120	21	Y58143	GLRaV-3 polyprotei
8	323	2.8	598	21	Y91963	PMWaV-1 helicase.
9	323	2.8	1128	11	R05107	Sequence encoded b
10	309	2.7	1116	15	R60608	Tobamovirus replic
11	290	2.5	2161	20	W87724	Rupestris stem pit
12	274.5	2.4	992	11	R06531	Protein encoded by
13	233	2.0	1704	15	R49657	Sequence of Heliot
14	185	1.6	2039	19	W56322	Haemophilus paraga
15	184.5	1.6	1693	19	W76368	Hepatitis E virus
16	184	1.6	1698	20	Y31381	HEV-US1 ORF1 prote
17	179.5	1.6	1693	15	R51264	HEV strain protein
18	179.5	1.6	1693	17	R91813	Hepatitis E virus
19	179.5	1.6	1693	19	W81519	Hepatitis E virus
20	176	1.5	1708	20	Y31384	HEV-US2 ORF1 prote
21	174	1.5	1708	20	W93408	Swine HEV ORF 1 pr
22	169.5	1.5	1693	19	W80196	Protein encoded by
23	169.5	1.5	1693	19	W71209	Protein encoded by
24	160.5	1.4	1693	12	R14618	Protein encoded by
25	158	1.4	1946	19	W47273	Lactobacillus bulg
26	154.5	1.3	1522	10	P93357	Sequence of the ca
27	152.5	1.3	1705	10	P94365	Sequence of part o
28	151.5	1.3	1706	11	R08031	Adenyl cyclase fro
29	146.5	1.3	1194	20	Y37825	Chlamydia trachoma
30	143.5	1.2	1252	16	R80530	B. sphaericus SLP.
31	142.5	1.2	1981	19	W42634	Protein sequence t
32	141.5	1.2	2353	17	R99393	Haemophilus adhesi
33	140	1.2	800	20	Y08624	Human secreted pro
34	140	1.2	1233	19	W41935	Nudaurelia beta-li
35	140	1.2	1233	19	W34536	Nudaurelia beta vi
36	140	1.2	3647	11	R05041	Filamentous haemag
37	138.5	1.2	2032	20	Y00238	Enterococcus faeca
38	138.5	1.2	2032	20	Y00240	Enterococcus faeca
39	138.5	1.2	2032	20	Y00242	Enterococcus faeca
40	137	1.2	928	20	W88417	Chlamydia pneumoni
41	137	1.2	949	20	Y35060	Chlamydia pneumoni
42	137	1.2	3596	21	Y87407	Bordetella pertuss
43	136.5	1.2	1978	20	Y27230	Amino acid sequenc
44	136.5	1.2	2285	20	W98149	Bacillus subtilis
45	135.5	1.2	1302	16	R84741	Multidrug resistan

ALIGNMENTS

RESULT 1
Y58148
ID Y58148 standard; Protein; 2237 AA.
XX
AC Y58148;
XX
DT 07-MAR-2000 (first entry)
XX
DE GLRaV-3 polyprotein.
XX
KW Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW viral disease; yield loss; sugar content; inhibition; infection;
KW replication; polyprotein; domain; proteinase; methyltransferase;
KW helicase; RNA-dependent; RNA polymerase; untranslated region;
KW transgenic plant; component; resistant; truncation; deletion; antisense;
KW expression; detection; antibody.
XX
OS Grapevine leafroll-associated virus 3.
XX
FH Key Location/Qualifiers
FT Domain 85..204
FT /note= "Proteinase domain (Y58143)"
FT Domain 460..731
FT /note= "Methyltransferase domain (Y58144)"
FT Domain 1922..2212
FT /note= "Helicase domain (Y58145)"
XX
PN WO9955880-A1.
XX
PD 04-NOV-1999.
XX
PF 29-APR-1999; 99WO-US09307.
XX
PR 29-APR-1998; 98US-0083404.
XX

PA (CORR) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Ling K;
XX
DR WPI; 2000-062035/05.
DR N-PSDB; Z49208.
XX
PT Newly isolated grapevine leafroll virus protein or polypeptide useful
PT for producing transgenic plants conferring viral disease resistance -
XX
PS Claim 3; Fig 16; 84pp; English.
XX
CC This sequence represents the polyprotein encoded by open reading
CC frame (ORF) 1a from the grapevine leafroll-associated virus 3
CC (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC grapes are grown. Although not lethal, it causes yield losses and
CC reduction in sugar content. The virus encodes several proteins,
CC which may serve as targets for the inhibition of viral infection or
CC replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC 1b; and a protein of unspecified function (Y58147), encoded by ORF 1l.
CC Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC useful for the generation of transgenic plants and plant components.
CC Such transgenic plants may be resistant to viral disease, for example,
CC this property being conferred on the plants via the use of nucleotides
CC encoding truncated or internally deleted proteins, or via the use of
CC antisense nucleotides to inhibit viral gene expression. The nucleotides
CC may additionally be used in the detection of viral nucleic acids in a
CC tissue sample. The proteins can be used to generate antibodies which
CC may be used to detect GLRaV-3 in plant samples. The isolation of
CC GLRaV-3 facilitates the production of agents that reduce the risk of
CC infection or damage by the virus in vineyards.
XX
SQ Sequence 2237 AA;

Query Match 100.0%; Score 11520; DB 21; Length 2237;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MDYIRPLRVFSFPHVNNTLEYVRYNKANGDVGAFLTTMKFIGNVKLSDFTPRCAAMIYIG 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mdyirplrvfsfphvnntleyvrynkangdvgaflttmkfignvklsdftprcaamiyig 60

Qy      61 KLTKGVKRTFVPPPVKGFARQYAVVSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVF 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 kltkgvkrtfvpppvkgfarqyavvsgsvsalrgdgkkvlmeartstsatsdvsdfdvvf 120

Qy     121 EAVSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKK 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 eavsnallvvhyhrvvpyapvkreqpkpavkqdeqkpkrqashwavkptavgvhvplpkk 180

Qy     181 QEALEPAQSVPQQSLEEKAALTFGLFFSKGGGDESDAVILRKGKLFNRALNVPIDVKNTF 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 qealepaqsvpqqsleekaaltfglffskgggdesdavilrkgklfnralnvpidvkntf 240

Qy     241 VWAKIWDEASRRRGYFYVKDRAVKFFPIVRGRATIEDFIVNTAPGCDVALPRIELWSMRE 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 vwakiwdeasrrrgyfyvkdravkffpivrgratiedfivntapgcdvalprielwsmre 300

Qy     301 RAFVCTTKGWCWFNNERLRGEIYRRRCFSSSFSIGFLMHLGFRSLKVIRFAGTNILHMPS 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 rafvcttkgwcwfnnerlrgeiyrrrcfsssfsigflmhlgfrslkvirfagtnilhmps 360

Qy     361 LNEERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGEILASVANALNQEEVYSSVVSSITNR 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 lneertfgwkggdvylpnvpktaivagdrtrlggeilasvanalnqeevyssvvssitnr 420

Qy     421 LVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPREREKLRELFPELSIQFS 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 lvlrdqsallshldtklcdmfsqrdamirekpshrcdvflkprereklrelfpelsiqfs 480

Qy     481 DSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDV 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 dsvrsshpfanamrscfngifsrrcgnvcffdiggsftyhvkaghvnchvcnpvldvkdv 540

Qy     541 KRRINEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCDSRADAGFMVDVYDISPQ 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 krrineilflstaggdsyvssdlltеaasksvsycsresqncdsradagfmvdvydispq 600

Qy     601 QVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEH 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 qvaeamdkkgalvfdialmfpvellygngevyleeldtlvkregdylaynvgqcgemyeh 660

Qy     661 SFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVG 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 sfsnvsgfftfsyvrtssgnvfkleyegyrcgyhhltmcraqkspgtevtyrslvpsfvg 720

Qy     721 KSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVI 780
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 kslvfipvvagssvsfktivldsdfvdriysyalntigtfenrtfeyavgavrsqkthvi 780

Qy     781 TGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTV 840
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 tgsrvvhskvdispddmwglvvavmaqaikdraksirsynfikasegslagvfklffqtv 840

Qy     841 GDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAF 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 gdcfsnavsvyakamvhdnfnvletlmsmprafirkvpgsvvvtictsgasdrlelrgaf 900

Qy     901 DISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNV 960
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 disketfgrklknsrlrvfsraivedsikvmkamktedgkplpitedsvyafimgnvsnv 960

Qy     961 HCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDAL 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 hctraglggskatvvssvskglvargaatkafsgitsffstgslfydrgltederldal 1020

Qy    1021 VRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGA 1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 vrtenainspvgiletsrvavskvvagtkefwsevslndfttfvlrnkvligifvaslga 1080

Qy    1081 APIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAV 1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 apiawkyrrgiaanarryagssyetlsslssqaagglrgltsstvsggslvvrrgfssav 1140

Qy    1141 TVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANT 1200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 tvtratvakrqvplallsfstsyaisgcsmlgiwahalprhlmfffglgtllgarasant 1200

Qy    1201 WKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNV 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 wkfggfsnnwcavpevvwrgksvssllpitlgvsliirgllndtipqlayvppvegrnv 1260

Qy    1261 YDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVT 1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ydetlryyrdfdydegagpsgtqheavpgddndgstssvssydvvtnvrdvgistngevt 1320

Qy    1321 GEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQS 1380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 geeethsprsvqytyveeevapsaavaerqgdpsgsgtadamafvesvkkgvddvfhqqs 1380

Qy    1381 SGETAREVEVDGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQA 1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 sgetarevevdgkgllpesvvgeaptqergraadgntaqtavnegdrepvqsslvsspqa 1440

Qy    1441 DIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEK 1500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 dipkvtqsevhaqkevkqevplatvsgatpivdekpapsvttrgvkiidkgkavahvaek 1500

Qy    1501 KQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKG 1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 kqvqveqpkqrsltinegkagkqlcmfrtcscgvqldvyneatiatrfsnaftfvdnlkg 1560
```

```
Qy   1561 RSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHA 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 rsavffsklgegytynggshvssgwpraleditaikypsvfdhclvqkykmgggvpfha 1620

Qy   1621 DDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNS 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ddeecypsdnpiltvnlvgkanfstkcrkggkvmvinvasgdyflmpcgfqrthlhsvns 1680

Qy   1681 IDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKAL 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 idegrisltfratrrvfgvgrmlqlaggvsdekspgvpnqqpqsqgatrtitpksggkal 1740

Qy   1741 SEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVV 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 segsgrevkgrstysiwceqdyvrkcewlradnpvmalepdytpmtfevvktgtsedavv 1800

Qy   1801 EYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQ 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 eylkylaigiertyrallmarniavttaegvlkvpnqvyeslpgfhvyksgtdlifhstq 1860

Qy   1861 DGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 dglrvrdlpyvliaekgiftkgkdvdavvalgdnlfvcddilvfhdainligalkvarcg 1920

Qy   1921 MVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPN 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 mvgesfksfeykcynappgggkttt1vdefvkspnstatitanvgssedinmavkkrdpn 1980

Qy   1981 LEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQI 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 leglnsattvnsrvvnfivrgmykrvlvdevhmmhqgllqlgvfatgaseglffgdinqi 2040

Qy   2041 PFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGK 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 pfinrekvfrmdcavfvpkkesvvytsksyrcpldvcyllssmtvrgtekcypekvvsgk 2100

Qy   2101 DKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGK 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 dkpvvrslskrpigttddvaeinadvylcmtqleksdmkrslkgkgketpvmtvheaqgk 2160

Qy   2161 TFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSV 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 tfsdvvlfrtkkaddslftkqphilvglsrhtrslvyaalsskldd kvgtyisdaspqsv 2220

Qy   2221 SDALLHTFAPAGCFRGI 2237
          |||||||||||||||||
Db   2221 sdallhtfapagcfrgi 2237
```

```
RESULT   2
W21636
ID   W21636 standard; Protein; 1390 AA.
XX
AC   W21636;
XX
DT   19-SEP-1997  (first entry)
XX
DE   Grapevine leafroll virus helicase.
XX
KW   GLRaV; grapevine; Vitis; rootstock; leafroll; disease resistance;
KW   transgenic plant; tristeza virus; citrus; helicase.
XX
OS   Grapevine leafroll associated virus type 3 isolate NY1.
XX
PN   WO9722700-A2.
XX
PD   26-JUN-1997.
XX
PF   20-DEC-1996;   96WO-US20747.
XX
PR   21-DEC-1995;   95US-0009008.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 1997-341691/31.
DR   N-PSDB; T72214.
XX
PT   DNA encoding grape-vine leaf-roll virus proteins - useful to impart
PT   viral-resistance to Vitis scion or root-stock cultivar(s)
XX
PS   Claim 4; Page 69-75; 172pp; English.
XX
CC   A helicase (W21636) was identified as the expression product of
CC   open reading frame 1a (T72215) of a double-stranded cDNA molecule
CC   derived from grapevine leafroll associated virus type 3 (GLRaV-3)
CC   isolate NY-1 dsRNA.  The helicase, and other GLRaV-3 polypeptides
CC   (see also W21637-47), can be expressed in transgenic plants as a
CC   means of imparting GLRaV resistance to Vitis scion or rootstock
CC   cultivars or tristeza virus resistance to citrus scion or rootstock
CC   cultivars, or used to raise antibodies useful for detection of
CC   GLRaV.
XX
SQ   Sequence   1390 AA;
```

```
  Query Match             45.5%;  Score 5247;  DB 18;  Length 1390;
  Best Local Similarity   74.1%;  Pred. No. 0;
  Matches 1055;  Conservative   96;  Mismatches  206;  Indels   66;  Gaps   17;

Qy    848 VSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETF 907
          || |||:::||||:||||:::|::|| |||||:|:|  ||| ||:|||||||:||: |
Db      1 vstyaksvmndnfniletlvtlpksfivkvpgsvlvsittsgisdklelrgafdvskknf 60

Qy    908 GRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGL 967
           |:|::|||||||||||||:||||| ||:|||||||| |||||||: ||:||||||||||
Db     61 srrlrssrlrvfsraivedtikvmkgmksedgkplpiaedsvyafmtgnmsnvhctragl 120

Qy    968 LGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAI 1027
          ||||||   |   ||  :|   || |||:||| | | |||| |||  |||||| | |:|:
Db    121 lggskacaaslavkgaasratgtklfsgltsflsagglfydegltpgerldaltrrehav 180

Qy   1028 NSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKY 1087
          |||||:||       :||:||| | ||:|| ||||||::|:||||:|  |:   |: |||
Db    181 nspvgllepgasvakrvvsgtkaflselsledfttfviknrvligvftlsmaltpvvwky 240

Qy   1088 RRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATV 1147
          || ||        :    :::  |   | | |      |   | | | |||| ||| ||
Db    241 rrniartgvdvfhrarsgtaaiglqclsggrslagdaargaltvtrgglssavavtrntv 300

Qy   1148 AKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFS 1207
          |:|||||||||||||:|||::|||||||||||||||||||| |  || |:|  ||::
Db    301 arrqvplallsfstsyavsgctllgiwahalprhlmfffglgtlfgvsastnswslggyt 360

Qy   1208 NNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRY 1267
          |:   |||: | |:|  |||    ||:||::||||::|:||| ||||:|||||||: | :
Db    361 nslftvpeltwegrsyrsllpqaalgislvvrgllsetvpqltyvppiegrnvydqalnf 420

Qy   1268 YRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNG--------EV 1319
          |||||||:|||||||  :: ||  |   |||| |        | |:  :|
Db    421 yrdfdyddgagpsgtagqsdpg-tntsdtssvfs--------ddglpasgggfdarveag 471

Qy   1320 TGEEETHSPR-SVQYTYVEE-EVAPSAAVAERQGD---PSGSGTADA------MAFVESV 1368
                ||| ||:: | |  :  |:   || : |   | |:   ::           :|
Db    472 pshavdesprgsvefvyrervdehpacgeaevekdlitplgtavlesppvgpeagsapnv 531

Qy   1369 KKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDRE 1428
          : |  :|        |   || ||    :| |   | | ||   : :|   | |  |   |
Db    532 edgcpev-----eaekcsevivd----vpss---eppvqevlestng--vqaarte---e 574

Qy   1429 PVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVP-----LATVSGA--------TPIVDEK 1475
           ||     :       | :||| :|:    :||      |   |||         |:  ||
Db    575 vvqgdtcga------gvaksev-sqrvfpaqvpaheagleassgavveplqvsvpvavek 627
```

```
Qy   1476 PAPSV-TTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGV 1534
Db    628 tvlsvekarelkavdkgkavvhakevknvpvktlprgalkisedtvrkelcmfrtcscgv 687

Qy   1535 QLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILT 1594
Db    688 qldvyneatiatrfsnaftfvdslkgrsavffsklgegytynggshvssgwpraledilt 747

Qy   1595 AIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVM 1654
Db    748 aikypsvfdhclvqkykmgggvpfhaddeecypsdnpiltvnlvgkanfstkcrkggkvm 807

Qy   1655 VINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKS 1714
Db    808 vinvasgdyflmpcgfqrthlhsvnsidegrisltfratrrvfgvgrmlqlaggvsdeks 867

Qy   1715 PGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNP 1774
Db    868 pgvpnqqpqsqgatrtitpksggkalsegsgrevkgrstysiwceqdyvrkcewlradnp 927

Qy   1775 VMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKV 1834
Db    928 vmalkpgytpmtfevvkagtsedavveylkylaigigrtyrallmarniavttaegvlkv 987

Qy   1835 PNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDN 1894
Db    988 pnqvyeslpgfhvyksgtdlifhstqdglrvrdlpyvfiaekgifikgkdvdavvalgdn 1047

Qy   1895 LFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSP 1954
Db   1048 lsvcddilvfhdainlmgalkvarcgmvgesfksfeykcynappgggkttmlvdefvksp 1107

Qy   1955 NSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMM 2014
Db   1108 nstatitanvgssedinmavkkrdpnleglnsattvnsrvvnfivrgmykrvlvdevymm 1167

Qy   2015 HQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPL 2074
Db   1168 hqgllqlgvfatgaseglffgdinqipfinrekvfrmdcavfvpkkesvvytsksyrcpl 1227

Qy   2075 DVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLE 2134
Db   1228 dvcyllssmtvrgtekcypekvvsgkdkpvvrslskrpigttddvaeinadvylcmtqle 1287

Qy   2135 KSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRS 2194
Db   1288 ksdmkrslkgkgketpvmtvheaqgktfsdvvlfrtkkaddslftkqphilvglsrhtrs 1347

Qy   2195 LVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAGCFRGI 2237
Db   1348 lvyaalsselddkvgtyisdaspqsvsdallhtfapagcfrgi 1390
```

```
RESULT   3
Y58145
ID   Y58145 standard; Protein; 291 AA.
XX
AC   Y58145;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein helicase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49205.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 15; Fig 11; 84pp; English.
XX
CC   This sequence represents the helicase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   291 AA;
```

```
  Query Match             13.0%;  Score 1495;  DB 21;  Length 291;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-113;
  Matches  291;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNL 1981
Db      1 vgesfksfeykcynappgggkttt1vdefvkspnstatitanvgssedinmavkkrdpnl 60

Qy   1982 EGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIP 2041
Db     61 eglnsattvnsrvvnfivrgmykrvlvdevhmmhqgllqlgvfatgaseglffgdinqip 120

Qy   2042 FINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGIEKCYPEKVVSGKD 2101
Db    121 finrekvfrmdcavfvpkkesvvytsksyrcpldvcyllssmtvrgtekcypekvvsgkd 180

Qy   2102 KPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKT 2161
Db    181 kpvvrslskrpigttddvaeinadvylcmtqleksdmkrslkgkgketpvmtvheaqgkt 240

Qy   2162 FSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYI 2212
Db    241 fsdvvlfrtkkaddslftkqphilvglsrhtrslvyaalsskldkvgtyi 291
```

```
RESULT   4
Y58144
```

```
ID   Y58144 standard; Protein; 272 AA.
XX
AC   Y58144;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein methyltransferase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49204.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 5; Fig 8; 84pp; English.
XX
CC   This sequence represents the methyltransferase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   272 AA;


  Query Match             12.4%;  Score 1432;  DB 21;  Length 272;
  Best Local Similarity   100.0%;  Pred. No. 5.9e-108;
  Matches  272;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     460 LKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTY 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 lkprereklrelfpelsiqfsdsvrsshpfanamrscfngifsrrcgnvcffdiggsfty 60

Qy     520 HVKAGHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRES 579
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 hvkaghvnchvcnpvldvkdvkrrineilflstaggdsyvssdllteaasksvsycsres 120

Qy     580 QNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEELDTL 639
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 qncdsradagfmvdvydispqqvaeamdkkgalvfdialmfpvellygngevyleeldtl 180

Qy     640 VKREGDYLAYNVGQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMC 699
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 vkregdylaynvgqcgemyehsfsnvsgfftfsyvrtssgnvfkleyegyrcgyhhltmc 240

Qy     700 RAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAG 731
           ||||||||||||||||||||||||||||||||
Db     241 raqkspgtevtyrslvpsfvgkslvfipvvag 272


RESULT   5
W73476
ID   W73476 standard; Protein; 2639 AA.
XX
AC   W73476;
XX
DT   29-MAR-1999  (first entry)
XX
DE   Grapevine leafroll virus type 2 (GRLaV-2) polyprotein.
XX
KW   GRLaV-2; closterovirus; grape; tobacco; transgenic plant;
KW   disease resistance; virus resistance; beet yellows virus;
KW   tristeza virus; protease; methyltransferase; helicase.
XX
OS   Grapevine leafroll virus type 2.
XX
PN   WO9853055-A1.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998;   98WO-US10313.
XX
PR   20-MAY-1997;   97US-0047194.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K,  Zhu H;
XX
DR   WPI; 1999-045307/04.
DR   N-PSDB; V08864 and V08874.
XX
PT   Grapevine leafroll virus (type 2) proteins and polypeptides - and
PT   encoding DNA, useful e.g. to impart grapevine leafroll resistance to
PT   grape and tobacco plants and detect grapevine leafroll virus
XX
PS   Claim 4; Page 25-33; 151pp; English.
XX
CC   This is the amino acid sequence of a 294 kDa polyprotein encoded by
CC   open reading frame ORF1a (see V08864) of grapevine leafroll virus
CC   type 2 (GLRaV-2) RNA (see V08874).  It contains conserved domains
CC   characteristic of 2 papain-like proteases, a methyltransferase and
CC   a helicase.  The GLTaV-2 genome includes 9 open reading frames
CC   (see V08864-72) encoding the polyprotein, an RNA-dependent RNA
CC   polymerase, heat shock proteins, coat proteins and other proteins
CC   of unknown function (see W73476-84).  These can be used to produce
CC   antibodies useful to detect GLRaV in samples e.g. by ELISA
CC   (claimed).  The nucleic acid molecules can be used to produce
CC   probes and primers for such detection, and to transform host cells
CC   (especially Agrobacterium vitis, Agrobacterium tumefaciens, grape,
CC   citrus, beet or tobacco cells) and produce transgenic plants
CC   (claimed).  They can be used to impart GLRaV-2 resistance to Vitis
CC   scion or rootstock cultivars or Nicotiana cultivars (claimed).
CC   Because extensive similarity exists between hsp70-related sequence
CC   regions of GLRaV-2 and other closteroviruses, the DNA may also be
CC   used to impart beet yellows virus resistance to beet cultivars or
CC   tristeza virus resistance to citrus scion cultivar/rootstock
CC   cultivars (claimed).
```

```
XX
SQ   Sequence   2639 AA;


  Query Match             6.9%;  Score 794.5;  DB 20;  Length 2639;
  Best Local Similarity   21.6%;  Pred. No. 2.2e-54;
  Matches  538;  Conservative  342;  Mismatches  913;  Indels  699;  Gaps  113;

Qy     152 QDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEP-AQSVPQQSL---------EEKAAL 201
Db     324 qssvktkredahrtvceraaggsveqprqkridekgcgrvpsggfshllvgnlnevrrkv 383

Qy     202 TFGLFFSKGGGDESDAVILRKGKLFNRALNVPIDVKNTFVWAK--------IWDEASRRR 253
Db     384 aagllrfrvggdmd---------fhrsfstqag-hrllvwrrssrsvclelyspsknfl 432

Qy     254 GYFYVK---DRAVKFFPIVRGRATIEDFIVNTAPGCDVALPRIELWSMRERAFVCTTKGW 310
Db     433 rydvlpcsgdyaamfsfaaggrfpl----------vlmtriry----------pngf 469

Qy     311 CWFNNERLRGEIYRRRCFSSSFSIG----------FLMHLGFRSLKVIRFAGTNILHMP 359
Db     470 cylahcryacafllrgfdpkrfdigafptaaklrnrmvselgersl---------- 515

Qy     360 SLNEERTFGWKG----------GDVYLPNVPKTAIVAGDRTRLGGEILASVANALNQEE 408
Db     516 glnlygaytsrgvfhcdydakfikdlrl----msaviag-kdgveevvpsditpamkqk- 569

Qy     409 VYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPREREKL 468
Db     570 ----tieavydrlyggtdsllklsiekdlidfkndvqslkkdrpivkvpfymseatqnsl 625

Qy     469 RELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKA-GHVN 527
Db     626 trfypqfelkfshsshsdhpaaaasrllenetlvrlcgn-svsdiggcplfhlhsktqrr 684

Qy     528 CHVCNPVLDVKDVKRR-INEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCDSRA 586
Db     685 vhvcrpvldgkdaqrrvvrdlqysnvrlgd----ddkile-gprnidichyplgacdhes 739

Qy     587 DAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEELDTLVKRE--G 644
Db     740 sammmvqvydaslyeicgamikkksrityltmvtpgefldgrecvymesldceievdvha 799

Qy     645 DYLAYNVGQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKS 704
Db     800 dvvmykfg--sscyshklsiikdimttpyl-tlggflfsvemyevrmgvnyfkitksevs 856

Qy     705 PGTEVT----YRSL--------VPSFVGKSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSY 752
Db     857 psisctkllryrransdvvkvklprfdkkrrmclp-------gydtiyldskfvsrvfdy 909

Qy     753 ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDR 812
Db     910 vvcncsavnsktfewvwsfikssksrviisgkiihkdvnldlkyvesiaavmlasgvrsr 969

Qy     813 ------AKSIRSYNFIKASEGSLAGVFKLFFQTVGDC-FSNAVSVYAKAMVHD---NFNV 862
Db     970 laseylaknlshfs-------------------gdcsfieatsfvlrekirnmtlnfn- 1008

Qy     863 LETLMSMPR--AFIRKVPGSVVVTICTSGASDRLELRGAFDISK------ETFGRKLKNS 914
Db    1009 -erllqlvkrvafatldvsfldldstlesitdfaeckvaieldelgclraeaenekirn- 1066

Qy     915 RLRVFSRAIVEDSIKVMKAMKTE---DGKPLPITEDSVYAFIMGNVSNVHC-----TRAG 966
Db    1067 --------lagdsiaaklaseivvdidskpspkq--------vgnsssenadkrevqrpg 1110

Qy     967 LLGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENA 1026
Db    1111 lrggsrngvvge----------------flhfvvdsalrlfkyatdqqriksyvr---f 1150

Qy    1027 INSPVGILETSRVAVSKVVAGTKEFWS----------------------EVSLNDFTT 1062

Db    1151 ldsavsfldynydnlsfilrvlsegyscmfaflanrgdlssrvrsavcavkevatscana 1210

Qy    1063 FVLRNKVLIGIFVASLGAAPIAWKYRR-GIAANARRYAG----------------SSYE 1104
Db    1211 svskakvmi-----tfaaavcammfnscgfsgdgreyksyihrytqvlfdtiffedssyl 1265

Qy    1105 TLSSLSSQAAGGLRGLTSS----TVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFS 1160
Db    1266 pievlssaicgaivtlfssgssislnafllqitkgfslevvv-------rnvvrvthgls 1318

Qy    1161 TSYAISGCSMLGIWAHALPRHLM-----------FFFGLGTLLGARASANTW-----KF 1203
Db    1319 tt-atdgv-irgvfsq-ivshllvgntgnvayqsafiagvvpllvkkcvslifilredty 1375

Qy    1204 GGFSNNWCAVPEVVW--------RGKSVSSLLLPITLGVSLIIRGL---------- 1241
Db    1376 sgfikh--gisefsflssilkflkgklvdel--------ksiiqgvfdsnkhvfkeatqe 1425

Qy    1242 -LNDTIPQ--LAYVPPVE---GRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDG- 1294
Db    1426 airttvmqvpvavvdalksaagkiynnftsr--rtfgkdegsssdgaceeyfscdegegp 1483

Qy    1295 STSSVSSYD---------VVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAA 1345
Db    1484 glkggssygfsilaffsrimwgarrlivkvkhecfgk---------lfefl------slk 1528

Qy    1346 VAERQGDPSGSGTADAMAFVESVKKGVDD-----VFHQQSSGETAREVEVDGKGLLPESV 1400
Db    1529 lhefrtrvfgknrtdv---------gvydflptgivetlssieecdqie----ellgddl 1575

Qy    1401 VGEAPTQERGRAADGNTAQTAVNEGD-REPVQSSLVSSPQADIPKVTQS---------- 1448
Db    1576 kg-----------dkdasltdmnyfefsedflasieeppfaglrggskniailaileyah 1624

Qy    1449 ---EVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQV 1505
Db    1625 nlfrivaskcskrplflafaelssaliek--fkevfpr------ksqlvaivreytq-rf 1675

Qy    1506 EQPKQRSLTINEGKAGKQLC------MFRTCSCGVQLDVYNE-------ATIATRFSNAF 1552
Db    1676 lrsrmralglnnefvvksfadllpalmkrkvsgsflasvyrplrgfsymcvsaerrekff 1735

Qy    1553 TFVDNLKGRSAVFFSK-----------------------LGEGYT---------- 1574
Db    1736 alv-cliglslpffvrivgakaceelvssarrfyerikiflrqkyvslsnffchlfssdv 1794

Qy    1575 --------YNGGSH---------------VSSGWPRALEDILTAIKYPSVFDHC--LVQK 1609
Db    1795 ddssasaglkggasrmtlfhllvrlasallslgw----eglkllshhnllflcfalvdd 1850

Qy    1610 ----YKMGGGVPFHADDEECYPSDNPILTV----------------NLVGKANFSTKCRK 1649
Db    1851 vnvlikvlgglsffv---------qpifslfaamllqpdrfveyseklvtafefflkcsp 1901

Qy    1650 GGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVF------GVGRML 1703
Db    1902 rapall------------kgf-------fecvanstvsktvrrllrcfvkmlklrkgrgl 1942

Qy    1704 QLAG-GVSDEKS-PGVPNQQPQSQGATRTITPKSGGKALS--------------EGSGRE 1747
Db    1943 radgrglhrqkavpvipsnrvvtdgverlsvkmqgvealrtelriledldsavieklnrr 2002

Qy    1748 VKGRSTYSIWCEQDYVRK-----------CEWLRADNPVMALE-----------PDYTPM 1785
Db    2003 -rnrdtn----ddeftrpaheqmqevttfcs--kansaglaleravlvedaikseklskt 2055

Qy    1786 TFEVVKTG--TSEDAVV-----EYLKYLAIGIER-----TYR----------ALLMARNI 1823
Db    2056 vnemvrkgsttseevavalsddeaveeisvaderddspktvriseylnrlnssfefpkpi 2115

Qy    1824 AVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHST----QDGLRV-----------RDL 1868
```

```
Db   2116 vvddnkdtggltnavre----fyymqelalfeihsklctyydqlrivnfdrsvapcseda 2171

Qy   1869 PYVLIAEKGIFTKGKDVDA-VVALGDNLFVCDDIL----------VFHDAINLIGALKV 1916
Db   2172 qlyvrkngstivqgkevrlhikdfhdhdflfdgkisinkrrrggnvlyhd--nl--afla 2227

Qy   1917 ARCGMVGESF--------KSFEYKCYNAPPGGGKTTTLVDEFV----KSPNSTATITANV 1964
Db   2228 snlflagypfsrsfvftnssvdillyeappgggktttlidsflkvfkkgevstmiltank 2287

Qy   1965 GSSEDINMAVKKRDPNLE---------GLNSATTVNSRVVNFIVRGMYKRVL-VDEVHMM 2014
Db   2288 ssqveilkkvekevsniecqkrkdkrspkksiytidaylmhh--rgcdadvlfidecfmv 2345

Qy   2015 HQGLLQLGVFATGASEGLFFGDINQIPFINR---EKVFRMDCAVFVPKKESVVYTSKSYR 2071
Db   2346 hagsvlacieftrchkvmifgdsrqihyierneldkclygdldrfv-dlqcrvygnisyr 2404

Qy   2072 CPLDVCYLLSSM------TVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINAD 2125
Db   2405 cpwdvcawlstvygnliatvkgese--------gks-----smrineinsvddlvpdvgs 2451

Qy   2126 VYLCMTQLEKSDMKRSLKGKG-KETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHI 2184
Db   2452 tflcmlqsekleiskhfirkgltklnvltvheaqgetyarvnlvrlkfqedepfksirhi 2511

Qy   2185 LVGLSRHTRSLVYAALSSKLDDKVGTYISDAS 2216
Db   2512 tvalsrhtdsltynvlaarrgdatcdaiqkaa 2543


RESULT   6
Y91969
ID   Y91969 standard; Protein; 309 AA.
XX
AC   Y91969;
XX
DT   19-JUL-2000  (first entry)
XX
DE   PMWaV-2 helicase.
XX
KW   Pineapple mealybug wilt virus 2; PMWaV-2; ORF 1a; helicase;
KW   transgenic pineapple; resistance; antiviral.
XX
OS   Pineapple mealybug wilt virus 2.
XX
PN   WO200017372-A2.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22152.
XX
PR   23-SEP-1998;   98US-0101461.
XX
PA   (UYHA-) UNIV HAWAII.
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Hu JS,  Karasev AV,  Dawson WO,  Melzer M;
XX
DR   WPI; 2000-283596/24.
DR   N-PSDB; A08690.
XX
PT   Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT   useful for protecting pineapples against the virus
XX
PS   Claim 17; Page 97-98; 112pp; English.
XX
CC   Pineapple mealybug wilt virus 2 (PMWaV-2), open reading frame (ORF) 1a
CC   is incomplete but encodes this protein which encompasses all motifs
CC   characteristic of viral helicases.  The DNA and protein sequences
CC   are useful for production of transgenic pineapple plant cultivars and
CC   also in transformation methods to impart resistance against the virus to
CC   pineapple plants.  Mealybug wilt is a major problem limiting profitable
CC   pineapple production in many pineapple growing areas worldwide, the
CC   present invention aims to overcome this problem.
XX
SQ   Sequence   309 AA;


  Query Match             5.5%;  Score 631.5;  DB 21;  Length 309;
  Best Local Similarity   44.7%;  Pred. No. 7.6e-43;
  Matches  136;  Conservative   50;  Mismatches   99;  Indels   19;  Gaps    4;

Qy   1949 EFVKSPNSTATITANVGSSEDINMAVKKRDP---------NLEGLNS------ATTVNSR 1993
Db      6 eyfkdrsgscimtanrgsaidindtiesidaanaskaasnnvsgvesidnyvcartvnsq 65

Qy   1994 VVNFIVRGM--YKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRM 2051
Db     66 imn--ckgvmnytcalvdemylmhkgllmlgvfssgarraifygdinqipfinrekcfys 123

Qy   2052 DCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKR 2111
Db    124 kegvycpgkdeiiytsesyrcpadvcmwssslkaqagsnrylkgvscnqrevvlrslskr 183

Qy   2112 PIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTK 2171
Db    184 pvvyaeqviqleadayitfkqeckekvvralravgrrdkvftsheaqgmtfgrvvlcrls 243

Qy   2172 KADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPA 2231
Db    244 atddsvfssephilvalsrhtqscvyatlsskladkvgaaidsvtrkevsdtvlktfvas 303

Qy   2232 GCFR 2235
Db    304 alfr 307


RESULT   7
Y58143
ID   Y58143 standard; Protein; 120 AA.
XX
AC   Y58143;
XX
DT   07-MAR-2000  (first entry)
XX
DE   GLRaV-3 polyprotein proteinase domain.
XX
KW   Grapevine; leafroll; grapevine leafroll-associated virus 3; GLRaV-3;
KW   viral disease; yield loss; sugar content; inhibition; infection;
KW   replication; polyprotein; domain; proteinase; methyltransferase;
KW   helicase; RNA-dependent; RNA polymerase; untranslated region;
KW   transgenic plant; component; resistant; truncation; deletion; antisense;
KW   expression; detection; antibody.
XX
OS   Grapevine leafroll-associated virus 3.
XX
PN   WO9955880-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-US09307.
XX
PR   29-APR-1998;   98US-0083404.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Gonsalves D,  Ling K;
XX
DR   WPI; 2000-062035/05.
DR   N-PSDB; Z49203.
XX
PT   Newly isolated grapevine leafroll virus protein or polypeptide useful
PT   for producing transgenic plants conferring viral disease resistance -
XX
PS   Claim 4; Fig 6; 84pp; English.
```

```
XX
CC   This sequence represents the proteinase domain of the
CC   polyprotein (Y58148) from the grapevine leafroll-associated virus 3
CC   (GLRaV-3). Leafroll is a serious viral disease, occurring wherever
CC   grapes are grown. Although not lethal, it causes yield losses and
CC   reduction in sugar content. The virus encodes several proteins,
CC   which may serve as targets for the inhibition of viral infection or
CC   replication. These proteins include the 242-248 kD polyprotein (Y58148),
CC   encoded by open reading frame (ORF) 1a, and which comprises a proteinase
CC   domain (Y58143), a methyltransferase domain (Y58144) and a helicase
CC   domain (Y58145); a RNA-dependent RNA polymerase (Y58146), encoded by ORF
CC   1b; and a protein of unspecified function (Y58147), encoded by ORF 11.
CC   Nucleotides encoding these proteins, or fragments thereof, and the 5'
CC   and 3' untranslated regions (UTRs) of the genome (Z49201-Z49202) are
CC   useful for the generation of transgenic plants and plant components.
CC   Such transgenic plants may be resistant to viral disease, for example,
CC   this property being conferred on the plants via the use of nucleotides
CC   encoding truncated or internally deleted proteins, or via the use of
CC   antisense nucleotides to inhibit viral gene expression. The nucleotides
CC   may additionally be used in the detection of viral nucleic acids in a
CC   tissue sample. The proteins can be used to generate antibodies which
CC   may be used to detect GLRaV-3 in plant samples. The isolation of
CC   GLRaV-3 facilitates the production of agents that reduce the risk of
CC   infection or damage by the virus in vineyards.
XX
SQ   Sequence   120 AA;


  Query Match             5.3%;  Score 605;  DB 21;  Length 120;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-41;
  Matches  120;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 VSGSVSALRGDGKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKRE 144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vsgsvsalrgdgkkvlmeartstsatsdvsdfdvvfeavsnallvvhyhrvvpyapvkre 60

Qy     145 QPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALTFG 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qpkpavkqdcqkpkrqashwavkptavgvhvplpkkqealepaqsvpqqsleekaaltfg 120


RESULT   8
Y91963
ID   Y91963 standard; Protein; 598 AA.
XX
AC   Y91963;
XX
DT   19-JUL-2000  (first entry)
XX
DE   PMWaV-1 helicase.
XX
KW   Pineapple mealybug wilt virus 1; PMWaV-1; ORF 1a; helicase;
KW   transgenic pineapple; resistance; antiviral.
XX
OS   Pineapple mealybug wilt virus 1.
XX
PN   WO200017372-A2.
XX
PD   30-MAR-2000.
XX
PF   22-SEP-1999;   99WO-US22152.
XX
PR   23-SEP-1998;   98US-0101461.
XX
PA   (UYHA-) UNIV HAWAII.
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Hu JS,  Karasev AV,  Dawson WO,  Melzer M;
XX
DR   WPI; 2000-283596/24.
DR   N-PSDB; A08684.
XX
PT   Isolated pineapple mealybug wilt virus proteins and polypeptides,
PT   useful for protecting pineapples against the virus
XX
PS   Claim 5; Page 79-82; 112pp; English.
XX
CC   Pineapple mealybug wilt virus 1 (PMWaV-1), open reading frame (ORF) 1a
CC   encodes this protein with all 8 motifs conserved in the so-called viral
CC   helicases.  The DNA sequences and related proteins are useful for
CC   production of transgenic pineapple plant cultivars and also in
CC   transformation methods to impart resistance against the virus to
CC   pineapple plants.  Mealybug wilt is a major problem limiting profitable
CC   pineapple production in many pineapple growing areas worldwide, the
CC   present invention aims to overcome this problem.
XX
SQ   Sequence   598 AA;


  Query Match             2.8%;  Score 323;  DB 21;  Length 598;
  Best Local Similarity   26.7%;  Pred. No. 2.9e-17;
  Matches  131;  Conservative   86;  Mismatches  175;  Indels   98;  Gaps   22;

Qy    1787 FEVVKT------GTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYE 1840
           |||:||      | |   :  :|         :|| |:      |: ::|   |   :
Db     117 feviktilantlgfsftgmnsfldi------esyrtlvr----rvsgsdgysavlealch 166

Qy    1841 SLPGFH--------VYKSGTDLIFH-------STQDGLRVRDLP------------YVL 1872
           || |           | |    || |       : |:  ||:  |             | |
Db     167 slhglraevdvldkvvkqeiplighkhelfcknvqeldrvkykpdgfhyynvelsniygl 226

Qy    1873 I-AEKGIFTKGKDVDAVVALGDNLFVCDDILVFHD-AINLIGAL-------KVARCGMVG 1923
           :   | ::      | ||   : : : :     |: : ||| :|       ||:    |
Db     227 vnthklvynn----dpvvkesegivllep----heisfnlirslalidllvkvs----ve 274

Qy    1924 ESFKSFE-YKCYNAPPGGGKTTTLVDEFV------KSPNSTATITANVGSSEDIN-MAVK 1975
           |  |: :  |  || || |||  :    :      |  ::   :|::  |:: :   | :
Db     275 evnkaidnvkfvnavpgagktyqikqrmlrwfdsekdgsallvltssrnsadtlkafaqe 334

Qy    1976 KRDPNLEGLNSATTVNSRVVNFIVRG----MYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
           ||   |  :    ||:: :  |  ||    :|| :|:|| :| | |:|:  : |    |
Db     335 kr---lgkmiqiltvdafl--fqargrnvrlyktllidecymthagilrgiiaavkpeec 389

Qy    2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKC 2091
           : :||  |:|||| |:   : :   |   :      : ||| |:|: :|:  |   :|
Db     390 vlygdrrqvpfinriklldnksflkpslgnysemlitrrcpadicwrmsn--vnngkk- 446

Qy    2092 YPEKVVSG------KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGK 2145
             ::: ||       : |||::|:: :     |       |  :  || ||:::      :
Db     447 -gdrlysgpvklftqskpvlksvtckafskgdhnlfsqvdrvmtftqnekneliseymsr 505

Qy    2146 G------KETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAA 2199
           |       :| : || |:||:|:  | |   |  || :|:  || || ||||| || |
Db     506 gigtiqdaktligtvaesqgetykrvhlvafkptddqvfssmphrlvalsrhtislqyfc 565

Qy    2200 LSSKLDDKVG 2209
           : :|::  :|
Db     566 ipnkmnkgig 575


RESULT   9
R05107
ID   R05107 standard; protein; 1128 AA.
XX
AC   R05107;
XX
DT   05-OCT-1990  (first entry)
XX
DE   Sequence encoded by RNA1 of AMV.
XX
KW   Alfalfa mosaic virus; AMV; cryptic infection; probe.
XX
OS   Nicotiniana tabacum L "Xanthi-nc".
XX
PN   US4921802-A.
XX
```

```
PD   01-MAY-1990.
XX
PF   13-JAN-1988;   88US-0144692.
XX
PR   05-MAY-1983;   83US-0492582.
PR   13-JAN-1988;   88US-0144692.
XX
PA   (PIOE ) PIONEER HI-BRED INTERNATIONAL LTD.
XX
PI   Hall TC,  Loesch-Fries SL,  Jarvis NP,  Barker RF;
XX
DR   WPI; 1990-171297/22.
DR   N-PSDB; Q04476.
XX
PT   Recombinant cDNA plasmids derived from alfalfa mosaic virus RNA -
PT   for detection of cryptic infection by alfalfa mosaic virus.
XX
PS   Disclosure; ; p; English.
XX
CC   The sequence is encoded by the bottom (B)- or "RNA1" genome fragment
CC   of AMV which is necessary to initiate infection.   DNA encoding the
CC   sequence can be used as a probe to detect cryptic infection of plants
CC   by AMV.
CC   See also R05108-09.
XX
SQ   Sequence   1128 AA;
```

Query Match 2.8%; Score 323; DB 11; Length 1128;
Best Local Similarity 17.0%; Pred. No. 9.1e-17;
Matches 319; Conservative 188; Mismatches 494; Indels 870; Gaps 65;

```
Qy    435 TKLCDMFSQ------RDAMIREKPSHRCDVF-----LKPREREKLRELFPELSIQFSDSV 483
Db     40 taigkvfseagrayaqdal----psdkgevlkisfsldatqqnilranfpgrrivfsnss 95

Qy    484 RSSHPFANAMRSCFNGIFSRRCGNV--CFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVK 541
Db     96 ssshcfaaahrlletdfvyrcfgntvdsiidlggnfvshmkvkrhnvhcccpildardga 155

Qy    542 RRINEILFLSTAGGDSYVSS--DLLTEAASKSVSYCSRESQNCDSRADAGFMV-DVYDIS 598
Db    156 rlterilslk-----syvrkhpeivgea-----dycmdtfqkcsrradyafaihptsdld 205

Qy    599 PQQVAEAMDKKGALVFDIALMFPVE---LLYGNGEVYLEELDTLVKREGDYLAYN-VGQC 654
Db    206 vgelacsldqkgvmkfictmmvdadadmlihsegeipnfnvrweidhkndlihfdfidep 265

Qy    655 GEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGT---EVTY 711
Db    266 nsgyshrfslikhylty--------navdlghaayr-------ierkqdfggvmvidlty 310

Qy    712 RSL-----VPSFVGKSLVFIPVVAGSSV-----------SFKTIVLDSDFVDRIYSYALN 755
Db    311 -slgfvpkmphsngrscawynrvkgqmvvhtvnegyyhhsyqtavrrkvlvdkkvltrvt 369

Qy    756 TIGTFENRTFEYAVGAVRSQKT-------HVITGSRVVHSKVDISPDDMWGLVVAVMAQA 808
Db    370 evafrqfrpnadahsaiqsiatmlssstnhtiiggvtlisgkplspddyipvattiyyrv 429

Qy    809 IK-----------DRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMV 856
Db    430 kklynaipemlslldkgerlstdavlkgsegpt--------wysgptflsald---kvnv 478

Qy    857 HDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRL 916
Db    479 pgdf-vakallslpk-----------------------------------rdlks--- 497

Qy    917 RVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVV 976
Db    498 -lfsrs-----------atshsertpvqde---------spvrc-------------- 520

Qy    977 SSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILET 1036
```

```
Db    521 ------------------------tdgvfy---------------------pirml-- 531

Qy   1037 SRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANAR 1096
Db    532 -----lkclgsdk--fesvtitd----------------------------------- 547

Qy   1097 RYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLAL 1156
Db    548 -------------------prsntettvd-----lyhsfqkki---------------- 566

Qy   1157 LSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEV 1216
Db    567 ------------------------etvfsfilgkidgp---------------------- 580

Qy   1217 VWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLR--YYRDFDYD 1274
Db    581 ----------------------------------splisdpvyfqsledvyyaew--- 601

Qy   1275 EGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYT 1334
Db    602 ----------------hqgnaidasny------------------------------- 612

Qy   1335 YVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKG 1394
Db    613 --------------------------------artllddirkqkeeslrakakevedaq 639

Qy   1395 LLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQK 1454
Db    640 klnrail---------------------------------------------qvhayl 652

Qy   1455 EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLT 1514
Db    653 eahpd----------------------grkieglglssqfiakipelaipapkplpef 688

Qy   1515 INEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYT 1574
Db    689 eknaetgeil----------rinphsdail-----eaidylkstsansiitlnklg---- 729

Qy   1575 YNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEE--CYPSDNPI 1632
Db    730 ---------------------------dhc--qwttkgldvvwagddkrrafipkkn-- 757

Qy   1633 LTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRA 1692
Db    758 ---twvgptarsyplakyeram-----skdgy--------------------------- 781

Qy   1693 TRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRS 1752
Db    782 ------------------------------------vtlrwdgevldanc---vrsls 800

Qy   1753 TYSI-WCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIE 1811
Db    801 qyeivfvdq----scvkasaeriipslek-------------------------algle 830

Qy   1812 RTYRALLMARNIAVTTAEGVL---KVPN--QVYESLPGFHVYKSGTDLIFHSTQDGLRVR 1866
Db    831 ---------ahfsvtivdgvagcgkttnikqiars----------------------- 856

Qy   1867 DLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESF 1926
Db    857 -------------sgqdvdli--ltsnrssadelketidcspltklhyirtc------- 893

Qy   1927 KSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNS 1986
Db    894 ---------------------dsy---------------------------------lms 899

Qy   1987 ATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINRE 2046
Db    900 asavka-----------qrlifdecflqhaglvyaaatlagcsevigfgdteqipfvsrn 948

Qy   2047 KVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKP--- 2103
```

```
Db    949 psfvfrhhkltgkverklit---wrspadatycl-------------ekyfyknkkpvkt 992

Qy   2104 ---VVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGK 2160
          |:||:   || :   |      :||| || ||: :|     || :  : | ||||||
Db    993 nsrvlrsievvpinspvsverntnalylchtqaekavlkaqthlkgcdn-ifttheaqgk 1051

Qy   2161 TFSDVVLFRTKKADDSLFT-KQP-----HILVGLSRHTRSLVYAALSSKLDDKVGTYISD 2214
          || :|   |  :   || | : |     | || |||| ::  |  ::   || :     |
Db   1052 tfdnvyfcrltrtstslttgrdpingpchglvalsrhkktfkyftiahdsddviynacrd 1111

Qy   2215 ASPQSVSDALL 2225
          |   :  |::|
Db   1112 ag--ntddsil 1120


RESULT 10
R60608
ID   R60608 standard; Protein; 1116 AA.
XX
AC   R60608;
XX
DT   21-JUN-1995  (first entry)
XX
DE   Tobamovirus replication enzyme subunit.
XX
KW   Tobamovirus; tobacco mosaic virus; vector; transcription; promoter;
KW   terminator; replication enzyme; replication site; ori; coat protein.
XX
OS   Tobacco mosaic virus.
XX
PN   JP06233688-A.
XX
PD   23-AUG-1994.
XX
PF   29-NOV-1993;   93JP-0297704.
XX
PR   01-DEC-1992;   92JP-0322100.
XX
PA   (SANY ) SANKYO CO LTD.
XX
DR   WPI; 1994-319190/40.
DR   N-PSDB; Q71546.
XX
PT   Improved transcriptional vector of plant virus - esp. vectors
PT   pL5013 and pL5020 which are prepd. and mass produced, then
PT   tested for stability and function
XX
PS   Disclosure; Page 17-19; 36pp; Japanese.
XX
CC   A transcriptional vector comprises in a 5'-3' direction (i) a 5'
CC   non-transcribed region of tobamovirus, (ii) a replicative enzyme of
CC   tobamovirus, (iii) a foreign gene connecting region, and (iv) a 3'
CC   non-transcribed region of tobamovirus.  The vector also has a
CC   transcriptional promoter region upstream of the tobamovirus
CC   region and a transcriptional terminator region downstream of the
CC   tobamovirus region, neither of which is recognised by the host cell,
CC   a replication initiation site which functions in the host cell, and
CC   a selective marker gene.  This sequence is a subunit of the
CC   tobamovirus replication enzyme described (See also R60609) and the
CC   sequence encoding it was used in a construct (See Q71546) which in
CC   turn was used in the construction of the transcriptional vectors
CC   pL5013 and pL5020.
XX
SQ   Sequence   1116 AA;


  Query Match             2.7%;  Score 309;  DB 15;  Length 1116;
  Best Local Similarity   16.8%;  Pred. No. 1.2e-15;
  Matches  315;  Conservative  205;  Mismatches  519;  Indels  840;  Gaps   64;

Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
          | |    |::: ::    | :  |     || : | |: ||     :|      :
Db      2 aytqtatssalletvrgnntlvndlakrrlydtav-defnard----rrpkvnfskvvse 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
          :    : :||  | | ::  : |  |  :||     :  :       :||||:|  |:
Db     57 eqtliatkaypefqitfyntqnavhslagglrsleleylmmqipygsltydiggnfashl 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVS----------SDLLTE 566
            |    | | | |||:|: |   :     |:||    |: :|           ::: |
Db    117 fkgrayvhccmpnldvrdimrhegqkdsielylslrlergnkhvpnfqkeafdryaempne 176

Qy    567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
              | || |   |  |  :  :|||   :   |: :|  |   |  |  ||
Db    177 vvchdtfqtcrhsqecytgrvyaialhsiydipadefgaallrknvhvcyaafhfsenll 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
          :  | |:|::   :|:|| | ::      : | ||:||:  :   :|   |:  |:
Db    237 ledshvnldeinacfqrdgdrltfsfasestlnyshsysnilkyvcktyfpasnrevymk 296

Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
          |:
Db    297 ef------------------------------------------------------------ 298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
            | |: ::    : | ||                    :    | |  ||
Db    299 lvtrvntwfckfsridtf------------------llykgvahkgvd----------- 328

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
                        |  | || | :       : :|:  | |  : :   : |:  |
Db    329 --------------seqfykamedawh-----ykktlamcnserilledsssvnywfpk 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
          :  :: :|                              |||| ||     | :|  |
Db    369 mrdmvivp----------------------------lfdislet----skrtrkevl--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
          : :| :  |:  ::|   | |  |  :| :|:    | |      :: |  |     | |
Db    394 vskdfvytvlnhirtyqakal--tysnvlsfvesirsrv------iingvtarsewdvdk 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041
            |:   : |        ||                          ::: : :|:   : :
Db    446 sllqslsmt--------ff-------------------------lhtklavlkdd-lli 470

Qy   1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097
          ||   | |      | |:||
Db    471 skfalgpktvsqhvwdeisl------------------------------------------ 490

Qy   1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157
                                              | :|    :  :  |:
Db    491 -------------------------------------afgnafpsikerlinrk------ 507

Qy   1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE-- 1215
                    :: |  :||                                  ||:
Db    508 ------------likitenale------------------------------irvpdly 524

Qy   1216 VVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDE 1275
          | :  : ||   : : :                    | :: |   :||   |
Db    525 vtfhdrlvseykmsvdm-------------------pvldirkkmeeteemy------- 557

Qy   1276 GAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTY 1335
                               :::|   |: |     :    ::         :|:
Db    558 ---------------------nalselsvlknsdkfdvdvfsqmc--------qsl---- 584

Qy   1336 VEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGL 1395
            || |  |             |  :  | | : |:   | | :    |  :
Db    585 ---evdpmta-------------akvivavmsnesgltltfeqpteanvalal------- 621

Qy   1396 LPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKE 1455
                    |:  :|:||                : :|:|   : | :  |
Db    622 -----------qdsekasdg---------------alvvtsrdveepsikgs------m 648

Qy   1456 VKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTI 1515
          : |: || :||       : |  | |                    :  ::|  :|  :
Db    649 argelqlaglsg------dvpessyt-----------------rseeiesleqfhmat 683
```

```
Qy   1516 NEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY 1575
Db    684 asslihkqmcsi----------vytgplkvqqmkn---fidslvaslsa---------- 719

Qy   1576 NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTV 1635
Db    720 ------------avsnlvkilkdtaaid--letrqkfg---------------------- 743

Qy   1636 NLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SIDEGR 1685
Db    744 ------------------vldvaskrwlvkpsaknhawgvvetharkyhvallehde-- 782

Qy   1686 ISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG 1745
Db    783 -----------fgi---------------------------------------------- 785

Qy   1746 REVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805
Db    786 ----------itcd-------nwrr---------------------vavssesvv----y 803

Qy   1806 LAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRV 1865
Db    804 sdmaklrtlrrllkdgephvssakvvl-------------------------------- 830

Qy   1866 RDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGES 1925
Db    831 ------------------vdgv------------------------------------ 834

Qy   1926 FKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------VKKRD 1978
Db    835 ------------pgcgk-------------tkeilsrvnfeedlilvpgrqaaemirrr 868

Qy   1979 PNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
Db    869 anasgiivatkdnvrtvdsflmnygkgarcqfkrlfideglmlhtgcvnflvemslcdia 928

Qy   2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTV 2085
Db    929 yvygdtqqipyinrvtgfpypahfaklevdevetrrttlrcpadvthflnqryeghvmct 988

Qy   2086 RGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
Db    989 ssekksvsqemvsgaasinpvskplkgk---------------iltftq---sd-keall 1029

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYAA--- 2199
Db   1030 srg-yadvhtvhevqgetyadvslvrltptpvsiiardsphvlvslsrhtkslkyytvvm 1088

Qy   2200 --LSSKLDD--KVGTYISD 2214
Db   1089 dplvsiirdlervssylld 1107
```

RESULT 11
W87724
ID W87724 standard; Protein; 2161 AA.
XX
AC W87724;
XX
DT 26-APR-1999 (first entry)
XX
DE Rupestris stem pitting associated virus RSPaV-1 replicase.
XX
KW RSPaV-1; grape; transgenic plant; disease resistance; replicase.
XX
OS Rupestris stem pitting associated virus.
XX
FH Key Location/Qualifiers
FT Misc-difference 1931
FT /note= "encoded by AG, apparent 1 nucleotide
FT deletion in open reading frame"
XX
PN WO9852964-A1.
XX
PD 26-NOV-1998.
XX
PF 20-MAY-1998; 98WO-US10391.
XX
PR 17-DEC-1997; 97US-0069902.
PR 20-MAY-1997; 97US-0047147.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Gonsalves D, Meng B;
XX
DR WPI; 1999-045297/04.
DR N-PSDB; V99284.
XX
PT Isolated proteins from Rupestris stem pitting-associated virus and
PT related nucleic acid - vectors, host cells and transgenic Vitis
PT cultivars that are resistant to the virus
XX
PS Claim 4; Page 18-25; 163pp; English.
XX
CC This is the amino acid sequence of the replicase of Rupestris stem
CC pitting associated virus RSPaV-1. It was deduced from the
CC nucleotide sequence of an open reading frame (ORF1) identified
CC in the RSPaV-1 genome (see V99284). Proteins and polypeptides
CC of RSPaV (see W87724-38), and nucleic acids encoding them (see
CC V99284-86) are claimed, as are expression systems, transformed
CC host cells, transgenic grape (Vitis) scion or rootstock cultivars
CC comprising DNA encoding a replicase, coat protein, or a protein
CC of a triple gene block, and a method of imparting RSPaV resistance
CC to Vitis scion or rootstock cultivars by transformation with such
CC DNA.
XX
SQ Sequence 2161 AA;

Query Match 2.5%; Score 290; DB 20; Length 2161;
Best Local Similarity 41.5%; Pred. No. 1.4e-13;
Matches 59; Conservative 26; Mismatches 53; Indels 4; Gaps 3;

```
Qy   1554 FVDNLKGRSAVFFSKLG-EGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKM 1612
Db    733 fkdrlrgrgasffskpgischsynggshtslgwpkfmdqilsstggrnyynsclaqiyee 792

Qy   1613 GGGVPFHADDEECYPSDNPILTVNLVGKANFS-TKCRK--GGKVMVINVASGDYFLMPCG 1669
Db    793 nsklalhkddescyeighkvltvnligsatftisksrnlvggnhcsltigpneffemprg 852

Qy   1670 FQRTHLHSVNSIDEGRISLTFR 1691
Db    853 mqcnyfhgvsnctpgrvsltfr 874
```

RESULT 12
R06531
ID R06531 standard; protein; 992 AA.
XX
AC R06531;
XX
DT 19-DEC-1990 (first entry)
XX
DE Protein encoded by RNA 1 gene of CMV.
XX
KW CMV; RNA 1.
XX
OS Cucumber mosaic virus.
XX
PN JP02167080-A.
XX
PD 27-JUN-1990.
XX
PF 19-DEC-1988; 88JP-0320015.
XX

```
PR   19-DEC-1988;   88JP-0320015.
XX
PA   (NOGY-) NOGYO SEIBUTSU IDEN.
XX
DR   WPI; 1990-241950/32.
DR   N-PSDB; Q05751.
XX
PT   Genome RNA 1 gene for prepn. of cucumber mosaic virus - codes
PT   protein contg. specified sequence of 994 amino acids used in prepn.
PT   of resistant plants.
XX
PS   Claim 1; Page 1; 10pp; Japanese.
XX
CC   The sequence was deduced from the DNA sequence which was prepd.
CC   from RNA isolated from CMV-O virus.  The gene can be used to
CC   prepare plants which are resistant to cucumber mosaic virus.
XX
SQ   Sequence   992 AA;


  Query Match             2.4%;  Score 274.5;  DB 11;  Length 992;
  Best Local Similarity   22.2%;  Pred. No. 6.3e-13;
  Matches  146;  Conservative  101;  Mismatches  267;  Indels  145;  Gaps   25;

Qy    1566 FSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEEC 1625
           : :||: |     | :|  |         |:    ||    |           ||
Db     434 yeqlgkmydmwnassisk-w-------faaltrrrvffssav----------halfptl 474

Qy    1626 YPSDNPILTVNLVGKANFSTKCR-KGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEG 1684
            | :    : |     |: :|   ||:   : ::|  |                :  :: :|
Db     475 rpreekefliklstfvtfneecsfdggeewdv-issaayv------------atqavtdg 521

Qy    1685 RISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGS 1744
           ::    :| :     :: |    |||  || ||:  |
Db     522 kvlaaqkaekl---aeklaqpvdevsd--spevpsstpd---------------------- 555

Qy    1745 GREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLK 1804
                  |  :  ::  | : : | |          :|:|  | :  |   |::||
Db     556 -------dtadvcgkeqevseldslsaqt--------rspit-rvaerat---amleyaa 596

Qy    1805 YLAIGIERTYRALLMARNIAVTTAEGVLKVPNQV-----YESLPG-----FHVYKSGTDL 1854
           |                : :  ||   : :: |         || |     |  | :   :
Db     597 y--------------ekqlhdttvsnlkriwnmaggddkrnslegnlkfvfdtyftvdpm 642

Qy    1855 IFHSTQDGLRVRDLPYVLIAEKGIFTKG---KDVDAVVALGDNLFVCDDILVFHDAINLI 1911
           :      |  :| :|  ::   |   :|   |    :  :     :|:      :::: :
Db     643 vnihfstgrwmrpvpegivysvgynerglgpksdgelfivnsecvicns-----eslsav 697

Qy    1912 GALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDIN 1971
                | : :          :   | |||| :   |   |::   :|||  |::|:
Db     698 trslqaptgtisq---------vdgvagcgkttaiksif--epstdmivtankksaqdvr 746

Qy    1972 MAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
           ||: |   : |      | :| ::|        |||||| ::| | |   :    |
Db     747 malfkssdskeacafvrtadsvllnec--ptvsrvlvdevvllhfgqlcavmsklkavra 804

Qy    2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTV----RG 2087
           : |||  || | :|:  | |  :  :| : |   |  ::| | ||  |:  |      :|
Db     805 icfgdseqiafssrdasfdmrfskiipdetsdadt--tfrspqdvvplvrlmatkalpkg 862

Qy    2088 TEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGK 2147
           |   | : |   |  | ||:: | | :   |   ::  |: ||| :|:    ||  : |
Db     863 thskytkwvsqsk---vkrsvtsrsiasvtlvdldssrfyitmtqadka----slisrak 915

Qy    2148 E---------TPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVY 2197
           |           : ||||:|| :   | | | |     || :  : || |:||  :  |
Db     916 emnlpktfwneriktvhesqgisedhvtlvrlkstkcdlfkqfsyclvaltrhkvtfry 974


RESULT  13
R49657
ID   R49657 standard; Protein; 1704 AA.
XX
AC   R49657;
XX
DT   12-SEP-1994  (first entry)
XX
DE   Sequence of Heliothis armigera replicase encoded by RNA 1.
XX
KW   HaSV; RNA 1; small RNA virus; replicase.
XX
OS   Heliothis armigera stunt virus.
XX
PN   WO9404660-A.
XX
PD   03-MAR-1994.
XX
PF   13-AUG-1993;   93WO-AU00411.
XX
PR   14-AUG-1992;   92AU-0004081.
PR   08-JUL-1993;   93US-0089372.
XX
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PA   (PACI-) PACIFIC SEEDS PTY LTD.
XX
PI   Christian PD,  Gordon KHJ,  Hanzlik TN;
XX
DR   WPI; 1994-083180/10.
DR   N-PSDB; Q58522.
XX
PT   Small RNA virus capable of infecting insect species, e.g.
PT   Heliothis - and transgenic plants contg. viral nucleic acid, for
PT   protection against insect pests
XX
PS   Disclosure; Figure 1; 183pp; English.
XX
CC   The inventors claim a virus comprising a genome hybridisable with
CC   the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC   those given in Figs 1 and 2 of the specification. Isolated proteins
CC   or polypeptide prepn. of the proteins or polypeptides derivable
CC   from the virus are also claimed.
CC   H. armigera larvae were raised and viral RNA was extracted. The virus
CC   RNAs were reverse transcribed into cDNA. Clone E3 represents 99.7%
CC   or RNA 1.(hr236 contains about 88% or RNA 2.)The full length clone
CC   of RNA 1 was completed using PCR. RNA 1 encodes a protein of mol.
CC   wt. 187,000 which is regarded as the replicase in view of its AA
CC   sequence similarity in certain limited regions to replicases of
CC   of other RNA viruses. The apparent mol. wt. of this protein upon
CC   in vitro translation of virus RNA and SDS-PAGE is 195,000. The
CC   sequence given in Figure 1 is in the same sense as the viral
CC   (positive-sense) RNA. There are other small ORFs at the 3' end,
CC   corresp. to the proteins P11a, P11b and P14.
XX
SQ   Sequence   1704 AA;


  Query Match             2.0%;  Score 233;  DB 15;  Length 1704;
  Best Local Similarity   20.9%;  Pred. No. 3.9e-09;
  Matches  206;  Conservative  108;  Mismatches  330;  Indels  342;  Gaps   47;

Qy    1425 GDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRG 1484
           |   | |  ::: |    :|  :    :  | :  |  ::  |  :     ||:    |
Db      54 gtlpptqhpilaghqrvaeevlhnfargrstvleigp--slhsalklhgapnapaadyhg 111

Qy    1485 VKIIDKGKAVAHVA--EKKQVQVEQPK---QRSLTINEGKAGKQLCMFRTCSC------G 1533
                     |:   | : |   :|:      ||  | |   :  |:    ||      |
Db     112 ctkygtrdgsrhitalesrsvatgrpefkadasllan-gigsrtfcvdgvgscafksrvg 170

Qy    1534 VQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDIL 1593
              :      : |:    :|||                        | | |:   :    |::|
Db     171 ianhslydvtl-eelanafe---------------------nhglhmvrafmhmpeell 207

Qy    1594 --------------TAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVG 1639
                           |: |     |  |    || :  |  : :            |
Db     208 ymdnvvnaelgyrfhvieepmavkdcafq----ggdlrlhfpeldfinesqerrierlaa 263
```

```
Qy   1640 KANFSTKCRKGGKVMVINVASGD----------------YFLM-----PCGF------QR 1672
          : ::| :           : |||                | |:     | ||      ||
Db    264 rgsysrra---------vifsgdddwgdaylhdfhtwlpyllvrnyptpfgfslhievqr 314

Qy   1673 THLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTIT 1732
           |  |:           | ||   |  ||| :           ||     |||  |
Db    315 rhgssie---------lritrappg-drmlav----------vpr---tsqglcr--i 347

Qy   1733 PKSGGKALSEGSGREVKGRSTYSIWCEQ------------------DYVRKCEWLRADN 1773
           |     |  : || | |     :|   |                       : ||
Db    348 pnifyya--dasgtehk-----tiltsqhkvnmllnfmqtrpekelvdmtvlmsfararl 400

Qy   1774 PVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLK 1833
            : :  : |  :: :         |  |: ::   |||   |      :|| ||:  :
Db    401 raivvasevtesswnispadlvrtvvslyvlhi---ierrraa------vavktakddvf 451

Qy   1834 VPNQVYESLPGFHVYKS--------GTDLIF-HSTQDGLRVRDLPYVL------------ 1872
              :|||   ||: |        |||::|     |  ||  |  ::
Db    452 getsfweslk--hvfgsccglrnlkgtdvvftkrvvdkyrvhslgdiicdvrlspeqvgf 509

Qy   1873 -------------------IAEKGIFT-------------KGKDVD-------------A 1887
                            : | | :               :| | |              |
Db    510 lpsrvpparvfhdreelealreagcynerpvpstppveepqgfdadlwhataaslpeyra 569

Qy   1888 VVALGDNLFVCDDILVFHDAINLIGALKVARC-GMVGESFKSFEYKCYNAPPGGGKTTTL 1946
           :  | |  |    :   :|:  |  | ::   |:          : |  ||| ||| ||
Db    570 tlqaglntdvkqlkitlenalktidgltlspvrgl----------emyegppgsgktgtl 619

Qy   1947 V----------------------DEFVKSPNSTATITANVGSSEDINMAVKKRDPNLEG 1983
           :                      |  :| |:::||        : : :|: :|    ||
Db    620 iaaleaaggkalyvaptrelreamdrrikppsasat--------qhvalailrr-ataeg 670

Qy   1984 LNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFI 2043
             || |           |:  | |  || :           : :|  :  ||::|| ||
Db    671 apfatvvidecf------mfplvyvaivhal----------spssrivlvgdvhqigfi 713

Qy   2044 NREKVFRMDCAVFVPKKESVVYTSKSYRCPL--DVCYLLSSMTVRGTEKCYPEKVVSGKD 2101
           :                    :   |   ||  ||       |   |::| |  ||:
Db    714 d--------------------fqgtsanmplvrdvvkqcrrrtfnqtkrc-padvva--- 749

Qy   2102 KPVVRSLSKRPIGTTDDVAEIN--ADVY-------LCMTQLEKSDMKRSLKGKGKETPVM 2152
           :||      |:  || |:  |  |       || || |||  :   :|       |
Db    750 ttffqslypgctttsgcvasishvapdyrnsqaqtlcftqeeks--rhgaeg------am 801

Qy   2153 TVHEAQGKTFSDVVL-FRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTY 2211
          |||||||:||: |:|  :    |:  |  :: |:|||::|||  |               |
Db    802 tvheaqgrtfasvilhyngstaeqkllaekshllvgitrhtnhl--------------y 846

Qy   2212 ISDASPQSVSDALLHTFAPAGCFRGI 2237
          | |  |    :  |:  | |  |  |
Db    847 ird--ptgdierqlnhsakaevftdi 870


RESULT 14
W56322
ID   W56322 standard; Protein; 2039 AA.
XX
AC   W56322;
XX
DT   19-AUG-1998  (first entry)
XX
DE   Haemophilus paragallinarum antigenic protein #2.
XX
KW   Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW   vaccine; chicken infectious coryza; CIC; fowl.
XX
OS   Haemophilus paragallinarum.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..70
FT                   /label= signal
FT   Protein         71..2039
FT                   /note= "antigenic protein"
XX
PN   WO9812331-A1.
XX
PD   26-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03222.
XX
PR   19-SEP-1996;   96JP-0271408.
XX
PA   (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
PA   (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.
XX
PI   Hamada F,  Matsuo K,  Sakaguchi M,  Tokiyoshi S,  Tokunaga E;
XX
DR   WPI; 1998-230318/20.
DR   N-PSDB; V22837.
XX
PT   Antigenic polypeptide from Haemophilus paragallinarum induces HI
PT   antibody production - and is useful for diagnosis of and preparation
PT   of vaccines for chicken infectious coryza
XX
PS   Claim 5; Page 71-87; 108pp; Japanese.
XX
CC   The present sequence represents an antigenic protein derived from
CC   Haemophilus paragallinarum strain C-53-47. The antigenic protein
CC   stimulates the production of HI antibodies in fowl. The protein
CC   and DNA coding for it can be used in the preparation of vaccines
CC   for the prevention of chicken infectious coryza (CIC). The protein
CC   and its antibodies can be used in the diagnosis and treatment of CIC.
XX
SQ   Sequence   2039 AA;


  Query Match             1.6%;  Score 185;  DB 19;  Length 2039;
  Best Local Similarity   20.3%;  Pred. No. 4.3e-05;
  Matches  263;  Conservative  147;  Mismatches  495;  Indels  388;  Gaps   61;

Qy    920 SRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSV 979
          ||   :||: :    |||           | :           :|  :|     : : |
Db    380 srtfakdsvai--gnkteasn----agsmaygy----------kakavgagaiaigaev 422

Qy    980 SKG-----------LVARGA-AT----------KAFSGITSFFSTGSLFYDRGLTEDERL 1017
          : |           |: ||| ||          ||   |  |    : |:|  ||:   |
Db    423 aagaefdssqagnlllnrgayatlksadksddikagdainvf----tqffdnmltqgshl 478

Qy   1018 DALVRTENA--INSPVGILETSRVAV-----SKVVAGTKEFWSEVSLNDFTTFVLRNKVL 1070
               ||   : :  | :: :  ||     : :  | | | |:           | |
Db    479 ----tyenttylttsagdikktlaavgdggknaiaignktfaska----------nsva 523

Qy   1071 IGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQ----AAGGLRGLTSSTVS 1126
          :|              |    | ||  :|  ||  :| |::       ||    ::| |
Db    524 lg-------------syalasaqna--falgsyslvsplaantivigvggyatgsnsfv- 567

Qy   1127 GGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFF 1186
          ||| |             :|:: | |   :| :| | :      | :  :|       |
Db    568 ggswv-------------stlsartv---vlgysasissdshdslamgvna-------fi 604

Qy   1187 GLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTI 1246
          | |                ||:  |:       |:  |   | :| :    |    ||
Db    605 gng-----------------snsslalgtgstiaknaks---pdslaigkdsridakdtd 644

Qy   1247 PQLAYVPPVEGRNVYDETLRYYRDFDYDE----------------------------- 1275
            : | |      ||||| | :| || ::
Db    645 ngvlytp-----qvydettrafrtfdenkdymrqamalgfnakvsrgkgkmetginsmai 699

Qy   1276 GAGPSGT-----------------QHEA----------VPGDDNDG--STSSVSSYDVV 1305
          ||    |               | ||          :|     |  | | | |   :
Db    700 garsqatlqnstalgvnaktdytweqleadpwvskgaisiptsgkigvisvgskgserri 759

Qy   1306 TNVRDVGISTNGEVTGEEETHSPR-----------SVQYTYVEEE----VAPSAAVAER 1349
            ||    : |:     : :|   |            |||  ||:      |   |   |
Db    760 vnvasgsldtdavnvaqlktieerfqseidllqngggvqylsvektningeagrvasqir 819
```

```
Qy   1350 QGDPSGS--GTADAMAFVESVKKGVDDVFHQQS---SGETAREVEVDGKGLLPESVVGEA 1404
Db    820 kgesykryvklktqllyldarkklngekfdqtsldkiskavqeleaeysgel------kt 873

Qy   1405 PTQERGRAADGNTAQTAVNE-GDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLA 1463
Db    874 taselnrvamqlnaettvndfgkfnqyktqienatnadsek------------------- 914

Qy   1464 TVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQ 1523
Db    915 nvgglspqv----iaqlkannnylndgakgqdsiafgwqaktse-------anglagkq 963

Qy   1524 LCMFRTCSCGVQLD--VYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHV 1581
Db    964 -----aiaigfqanssaenaisigtnsdtsmt------gavai-----gkgatvtaggkp 1007

Qy   1582 SSGWPRALEDILTAIKYPSVFDHCLVQKYK--MGGGVPFHADDEECYPSDNPILTVNLVG 1639
Db   1008 s-----------ialgqdstvansaisrtssvmingltf--nnfagspetlgvlsigtag 1054

Qy   1640 KANFSTKCRKGGKVMVINVASGDYF----------LMPCGFQRTHL-HSVNSIDEGRIS 1687
Db   1055 ke------rk-----ivnvaagdisqtsteaingsqlyatnfmlnklaqsvktnfggnan 1103

Qy   1688 LTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQ--SQGATRTITPKSGGKALS--EG 1743
Db   1104 latdqtitftnig------gtgqdtihdainnvltklislsateeeevvsgeavydalkg 1157

Qy   1744 SGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPD----------------YTP 1784
Db   1158 akptvsaeankgitglvdvvkk-----anspi-tvepstdnnkkktftvglmkdiegvns 1211

Qy   1785 MTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKV--------PN 1836
Db   1212 itid--ksqqdlnqvtgrmssagl----tfkkgdttngstttfaedgltidsttnsaqtn 1265

Qy   1837 QVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLF 1896
Db   1266 lvkvsrdgfsvkngsdesklastklsigaenaehvevtksgialk----------adnts 1315

Qy   1897 VCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTL-VDEFVKSPN 1955
Db   1316 dkssitlaqdaitlagnatgtaikltgvadgnitvnskdavnggqlrtllgvdsgakigg 1375

Qy   1956 STATITANVGSS-----EDINMAVKKRDPN------LEGLN--SATTVNSRVVNFIVRGM 2002
Db   1376 tekttiseaisdvkqaltdatlaykadnkngktvkltdglnftsttnidasved---ngv 1432

Qy   2003 YKRVLVDEVHMMHQGLLQLGVFATGASEGLFFG 2035
Db   1433 vkftlkdkl----tglktiateslnasqniiag 1461
```

RESULT 15
W76368
ID W76368 standard; Protein; 1693 AA.
XX
AC W76368;
XX
DT 03-DEC-1998 (first entry)
XX
DE Hepatitis E virus hollow particle protein #1.
XX
KW Hollow particle protein; virus; antibody; detection; immunoassay;
KW infection.
XX
OS Hepatitis virus.
XX
FH Key Location/Qualifiers
FT Protein 1..1693
FT /note= "Partial sequence"
XX
PN JP10234383-A.
XX
PD 08-SEP-1998.
XX
PF 28-FEB-1997; 97JP-0062445.
XX
PR 28-FEB-1997; 97JP-0062445.
XX
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
XX
DR WPI; 1998-535037/46.
DR N-PSDB; V61687.
XX
PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
XX
PS Claim 10; Page 17-24; 29pp; Japanese.
XX
CC This sequence represents a Hepatitis E viral hollow particle protein.
CC This polypeptides can be used to raise antibodies to detect HEV
CC infection in samples, e.g. by immuno-assay based techniques, and the
CC nucleic acid can be used for the same in nucleic acid hybridisation
CC assays. The polypeptides and nucleic acids allow more accurate
CC detection of HEV than previously possible.
XX
SQ Sequence 1693 AA;

Query Match 1.6%; Score 184.5; DB 19; Length 1693;
Best Local Similarity 22.7%; Pred. No. 3.4e-05;
Matches 131; Conservative 75; Mismatches 191; Indels 181; Gaps 30;

```
Qy   1715 PGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTY----SIWCEQDYVRKCEWLR 1770
Db    751 psapaldepasgat------agapaithqtarhrrllftypdgskvfagslfesactwl- 803

Qy   1771 ADNPVMALEPDYT-------------PMTFE----VVKTGTSEDAVV----------EY- 1802
Db    804 ----vnasnvdhrpggglchafyqrypasfdaasfvmrdgaaaytltprpiihavapdyr 859

Qy   1803 LKYLAIGIERTYRALLMARNIAV--TTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQ 1860
Db    860 lehnpkrleaayretcsrlgtaaypllgtgiyqvpig-----psfdawern-----hrpg 909

Qy   1861 DGLRVRDL------------PYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAI 1908
Db    910 delylpelaarwfeanrptrptltited------------vartanl-----------ai 946

Qy   1909 NLIGALKVARC--------GMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATI 1960
Db    947 eldsatdvgracagcrvtpgvv-------qyqftagvpgsgksrsi---------tqadv 990

Qy   1961 TANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQ 2020
Db    991 dvvvvptrelrnawrrr-----gfaaftphtaarvt-----qgrrvvideapslpphlll 1040

Qy   2021 LGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPK-KESVVYTS---KSYRCPLDV 2076
Db   1041 lhm--qraatvhllgdpnqipaidfeh------aglvpairpdlaptswwhvthrcpadv 1092

Qy   2077 CYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPIGTTDDVAEINADVYLCMTQLE 2134
Db   1093 cel-----irg---aypmiqttsr---vlrslfwgepavgqk----------lvftqaa 1130

Qy   2135 KSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRS 2194
Db   1131 kaanpgs----------vtvheaqgatytettiiatadargliqssrahaivaltrhtek 1180

Qy   2195 LVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAG 2232
Db   1181 sviidasgllre-vg----------isdaivnnfflag 1207
```

Search completed: January 31, 2001, 15:29:28
Job time: 217 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:26:27 ; Search time 27.64 Seconds
(without alignments)
1453.325 Million cell updates/sec

Title: US-09-301-906-15
Perfect score: 11520
Sequence: 1 MDYIRPLRVFSFPHVNNTLE.....QSVSDALLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5247	45.5	1390	2	US-08-770-544-2	Sequence 2, Appli
2	290	2.5	2161	3	US-09-081-320-3	Sequence 3, Appli
3	179.5	1.6	1693	3	US-08-840-316-1	Sequence 1, Appli
4	179.5	1.6	1693	4	PCT-US93-08849A-1	Sequence 1, Appli
5	179.5	1.6	1693	4	PCT-US93-08849-1	Sequence 1, Appli
6	169.5	1.5	1693	3	US-08-478-507-7	Sequence 7, Appli
7	150.5	1.3	1222	2	US-08-682-517-15	Sequence 15, Appl
8	150.5	1.3	1252	2	US-08-682-517-9	Sequence 9, Appli
9	149.5	1.3	1794	5	5183745-6	Patent No. 5183745
10	142.5	1.2	1489	5	5183745-2	Patent No. 5183745
11	142.5	1.2	1848	4	PCT-US95-10661A-6	Sequence 6, Appli
12	138.5	1.2	1912	1	US-08-409-995-4	Sequence 4, Appli
13	138.5	1.2	1912	3	US-08-685-467-4	Sequence 4, Appli
14	135.5	1.2	1302	1	US-08-232-537-2	Sequence 2, Appli
15	135	1.2	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
16	134.5	1.2	1541	4	PCT-US95-10661A-3	Sequence 3, Appli
17	134	1.2	1027	3	US-08-446-137B-2	Sequence 2, Appli
18	131.5	1.1	2756	1	US-08-375-709-11	Sequence 11, Appl
19	131.5	1.1	2756	1	US-08-752-929-11	Sequence 11, Appl
20	131.5	1.1	2756	3	US-09-090-793-7	Sequence 7, Appli
21	131	1.1	1611	2	US-08-804-227C-5	Sequence 5, Appli
22	130	1.1	2446	2	US-08-551-356-2	Sequence 2, Appli
23	130	1.1	2446	4	PCT-US93-12687-2	Sequence 2, Appli
24	129	1.1	4545	2	US-08-804-227C-14	Sequence 14, Appl
25	129	1.1	4550	2	US-08-804-227C-8	Sequence 8, Appli
26	129	1.1	4550	2	US-08-804-198-2	Sequence 2, Appli
27	127.5	1.1	1780	1	US-08-769-309A-5	Sequence 5, Appli
28	127.5	1.1	1780	3	US-08-994-570-5	Sequence 5, Appli
29	127	1.1	1026	1	US-08-194-290-7	Sequence 7, Appli
30	127	1.1	1026	2	US-08-614-377A-7	Sequence 7, Appli
31	125	1.1	2324	1	US-08-283-857-1	Sequence 1, Appli
32	125	1.1	2324	4	PCT-US95-09819-1	Sequence 1, Appli
33	125	1.1	2327	5	5455158-1	Patent No. 5455158
34	124.5	1.1	1612	1	US-08-169-927-2	Sequence 2, Appli
35	124.5	1.1	4472	2	US-08-804-227C-2	Sequence 2, Appli
36	124	1.1	1529	2	US-08-728-470-10	Sequence 10, Appl
37	124	1.1	1651	1	US-08-447-411-2	Sequence 2, Appli
38	124	1.1	2431	1	US-07-920-281C-2	Sequence 2, Appli
39	123.5	1.1	1069	1	US-07-777-715-9	Sequence 9, Appli
40	123.5	1.1	1069	1	US-08-170-126-4	Sequence 4, Appli
41	123.5	1.1	1069	3	US-08-954-418-4	Sequence 4, Appli
42	123.5	1.1	1545	4	PCT-US95-10661A-4	Sequence 4, Appli
43	123.5	1.1	2409	5	5180808-2	Patent No. 5180808
44	123	1.1	628	3	US-08-656-034-2	Sequence 2, Appli
45	122	1.1	1346	3	US-09-320-878-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-770-544-2
; Sequence 2, Application US/08770544
; Patent No. 5907085
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Ling, Kai-Shu
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
; TITLE OF INVENTION: THEIR USES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,544
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60009008
; FILING DATE: 21-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1390 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-770-544-2

Query Match 45.5%; Score 5247; DB 2; Length 1390;
Best Local Similarity 74.1%; Pred. No. 0;
Matches 1055; Conservative 96; Mismatches 206; Indels 66; Gaps 17;

Qy 848 VSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETF 907

```
Db      1 VSTYAKSVMNDNFNILETLVTLPKSFIVKVPGSVLVSITTSGISDKLELRGAFDVSKKNF 60

Qy    908 GRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGL 967
Db     61 SRRLRSSRLRVFSRAIVEDTIKVMKGMKSEDGKPLPIAEDSVYAFMTGNMSNVHCTRAGL 120

Qy    968 LGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAI 1027
Db    121 LGGSKACAASLAVKGAASRATGTKLFSGLTSFLSAGGLFYDEGLTPGERLDALTRREHAV 180

Qy   1028 NSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKY 1087
Db    181 NSPVGLLEPGASVAKRVVSGTKAFLSELSLEDFTTFVIKNRVLIGVFTLSMALTPVVWKY 240

Qy   1088 RRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATV 1147
Db    241 RRNIARTGVDVFHRARSGTAAIGLQCLSGGRSLAGDAARGALTVTRGGLSSAVAVTRNTV 300

Qy   1148 AKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFS 1207
Db    301 ARRQVPLALLSFSTSYAVSGCTLLGIWAHALPRHLMFFFGLGTLFGVSASTNSWSLGGYT 360

Qy   1208 NNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRY 1267
Db    361 NSLFTVPELTWEGRSYRSLLPQAALGISLVVRGLLSETVPQLTYVPPIEGRNVYDQALNF 420

Qy   1268 YRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNG--------EV 1319
Db    421 YRDFDYDDGAGPSGTAGQSDPG-TNTSDTSSVFS--------DDGLPASGGGFDARVEAG 471

Qy   1320 TGEEETHSPR-SVQYTYVEE-EVAPSAAVAERQGD---PSGSGTADA------MAFVESV 1368
Db    472 PSHAVDESPRGSVEFVYRERVDEHPACGEAEVEKDLITPLGTAVLESPPVGPEAGSAPNV 531

Qy   1369 KKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDRE 1428
Db    532 EDGCPEV-----EAEKCSEVIVD----VPSS---EPPVQEVLESTNG--VQAARTE---E 574

Qy   1429 PVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVP-----LATVSGA--------TPIVDEK 1475
Db    575 VVQGDTCGA------GVAKSEV-SQRVFPAQVPAHEAGLEASSGAVVEPLQVSVPVAVEK 627

Qy   1476 PAPSV-TTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGV 1534
Db    628 TVLSVEKARELKAVDKGKAVVHAKEVKNVPVKTLPRGALKISEDTVRKELCMFRTCSCGV 687

Qy   1535 QLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILT 1594
Db    688 QLDVYNEATIATRFSNAFTFVDSLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILT 747

Qy   1595 AIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVM 1654
Db    748 AIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVM 807

Qy   1655 VINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKS 1714
Db    808 VINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKS 867

Qy   1715 PGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNP 1774
Db    868 PGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNP 927

Qy   1775 VMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKV 1834
Db    928 VMALKPGYTPMTFEVVKAGTSEDAVVEYLKYLAIGIGRTYRALLMARNIAVTTAEGVLKV 987

Qy   1835 PNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDN 1894
Db    988 PNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVFIAEKGIFIKGKDVDAVVALGDN 1047

Qy   1895 LFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSP 1954
```

```
Db   1048 LSVCDDILVFHDAINLMGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTMLVDEFVKSP 1107

Qy   1955 NSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMM 2014
Db   1108 NSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVYMM 1167

Qy   2015 HQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPL 2074
Db   1168 HQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPL 1227

Qy   2075 DVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLE 2134
Db   1228 DVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLE 1287

Qy   2135 KSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRS 2194
Db   1288 KSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRS 1347

Qy   2195 LVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAGCFRGI 2237
Db   1348 LVYAALSSELDDKVGTYISDASPQSVSDALLHTFAPAGCFRGI 1390
```

RESULT 2
US-09-081-320-3
; Sequence 3, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,147
; FILING DATE: 20-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/069,902
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1722
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2161 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-081-320-3

Query Match 2.5%; Score 290; DB 3; Length 2161;

```
  Best Local Similarity   41.5%;  Pred. No. 8.5e-16;
  Matches   59;  Conservative   26;  Mismatches   53;  Indels    4;  Gaps    3;

Qy   1554 FVDNLKGRSAVFFSKLG-EGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKM 1612
          | | |:|| | |||| |   ::|||||| | |||: :: ||::    : :: || | |:
Db    733 FKDRLRGRGASFFSKPGISCHSYNGGSHTSLGWPKFMDQILSSTGGRNYYNSCLAQIYEE 792

Qy   1613 GGGVPFHADDEECYPSDNPILTVNLVGKANFS-TKCRK--GGKVMVINVASGDYFLMPCG 1669
             :  | ||| ||   : :|||||:| | |: :| |   ||    : :   ::| || |
Db    793 NSKLALHKDDESCYEIGHKVLTVNLIGSATFTISKSRNLVGGNHCSLTIGPNEFFEMPRG 852

Qy   1670 FQRTHLHSVNSIDEGRISLTFR 1691
             |  : | |::   ||:|||||
Db    853 MQCNYFHGVSNCTPGRVSLTFR 874


RESULT   3
US-08-840-316-1
; Sequence 1, Application US/08840316
; Patent No. 6054567
;  GENERAL INFORMATION:
;    APPLICANT:  Emerson, Suzanne U., Purcell, Robert H.,
;    APPLICANT:  Tsarev, Sergei. A., and Robinson, Robin A.
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  111
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/840,316
;      FILING DATE:  11-APR-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard W. Bork
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4255
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 AMINO ACID RESIDUES
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  UNKNOWN
;      TOPOLOGY:  UNKNOWN
US-08-840-316-1


  Query Match             1.6%;  Score 179.5;  DB 3;  Length 1693;
  Best Local Similarity   21.3%;  Pred. No. 4.4e-06;
  Matches  189;  Conservative   98;  Mismatches  320;  Indels  281;  Gaps   45;

Qy   1521 GKQLCMFRTC----SCGVQLD----VYNEA------TIATRFSNAFTFVDNLKGRSAVFF 1566
          |:||  :  |    | |  ||    |::|:      |   :  : |       |:    |
Db    425 GRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAIRKAVSKFCCFMKWLGQECTCF 484

Qy   1567 SKLGEGY---------TYNG-------------GSHVSSGWPRALEDILTAIKYPSVFD 1603
             :  ||           | |              ||:|  |   ||: :  |:  |:

Db    485 LQPAEGVVGDQGHDNEAYEGSDVDPAESAISDISGSYVVPG--TALQPLYQALDLPA--- 539

Qy   1604 HCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDY 1663
             : :  ::   |       :   | |    |:|   | |     | |:  |       .
Db    540 EIVARAGRLTATV-------KVSQVDGRIDCETLLGNKTFRTSF-VDGAVLETN------ 585

Qy   1664 FLMPCGFQRTHLH---SVNSIDEGRISLTFRATRR-------VFGVGRMLQLAGGVSDEK 1713
             | :| :|    | :::  |  |||: |:           |:      | |||
Db    586 -----GPERHNLSFDASQSTMAAGPFSLTYAASAAGLEVRYVAAGLDHRAVFAPGVSPRS 640

Qy   1714 SPGVP----------NQQPQ-----------SQGATRTITPKSGGKALSEGSGREVKGRS 1752
          :||            |:: |           :|      | | |      |     | |
Db    641 APGEVTAFCSALYRFNREAQRLSLTGNFWFHPEGLLGPFAPFSPGHVWE--SANPFCGES 698

Qy   1753 TY--SIWCEQDYVRKCEW-----------LRADNPVMAL-----EPDYTPMTFEVVKTG 1793
          |     | | | |                  ||  |  |       || :| |      |
Db    699 TLYTRTWSEVDAVPSPAQPDLGFTSEPSIPSRAATPTPAAPLPPPAPDPSP-TLSAPARG 757

Qy   1794 TSEDAVVEYLKYLAIGIERTYRAL--------LMARNIAVTTAEGVLKVPN--------- 1836
                  :     | | |        : | ::  :|   ::   |
Db    758 EPAPGATARAPAITHQTARHRRLLFTYPDGSKVFAGSLFESTCTWLVNASNVDHRPGGGL 817

Qy   1837 --QVYESLP-----------GFHVYKSGTDLIFHSTQDGLRVRDLP------------- 1869
             |:  |             |   |      | |:     |:   |
Db    818 CHAFYQRYPASFDAASFVMRDGAAAYTLTPRPIIHAVAPDYRLEHNPKRLEAAYRETCSR 877

Qy   1870 -----YVLIAEKGIFTK--GKDVDAVV---ALGDNLFV-------------CDDILVFH 1905
             | |:   ||:    |   ||       || |::                |  : :
Db    878 LGTAAYPLLG-TGIYQVPIGPSFDAWERNHRPGDELYLPELAARWFEANRPTCPTLTITE 936

Qy   1906 D-------AINLIGALKVARC--------GMVGESFKSFEYKCYNAPPGGGKTTTLVDEF 1950
          |       || |  |  | |          |:|       :|:     || ||: ::
Db    937 DVARTANLAIELDSATDVGRACAGCRVTPGVV-------QYQFTAGVPGSGKSRSI---- 985

Qy   1951 VKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDE 2010
             : | :   |  : ::  | ::|     |  : |   :  |        :||::||
Db    986 -----TQADVDVVVVPTRELRNAWRRR-----GFAAFTPHTAARVT-----QGRRVVIDE 1030

Qy   2011 VHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPK-KESVVYTS-- 2067
             :   || | :    |:     || |||| |: |       |  ||  :  :  ||
Db    1031 APSLPPHLLLLHM--QRAATVHLLGDPNQIPAIDFEH------AGLVPAIRPDLAPTSWW 1082

Qy   2068 -KSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPIGTTDDVAEINA 2124
             ::||| ||| |     :||    ||   : :   |:|||   :  :|
Db    1083 HVTHRCPADVCEL-----IRG---AYPMIQTTSR---VLRSLFWGEPAVGQK-------- 1123

Qy   2125 DVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHI 2184
             |  ||  |:    |          :|||||| |:::  :  |  |   : : : |
Db    1124 ---LVFTQAAKAANPGS----------VTVHEAQGATYTETTIIATADARGLIQSSRAHA 1170

Qy   2185 LVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAG 2232
          :| |:|||   |       | : ||          :|||::: |  ||
Db    1171 IVALTRHTEKCVIIDAPGLLRE-VG----------ISDAIVNNFFLAG 1207


RESULT   4
PCT-US93-08849A-1
; Sequence 1, Application PC/TUS9308849A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
```

```
;       COMPUTER:  IBM PC COMPATIBLE
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  WORDPERFECT 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US93/08849A
;       FILING DATE:  17-SEP-1993
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US07/947,263
;       FILING DATE:  18-SEP-1992
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  William S. Feiler
;       REGISTRATION NUMBER:  26,728
;       REFERENCE/DOCKET NUMBER:  2026-4032 PCT
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 758-4800
;       TELEFAX:  (212) 751-6849
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1693 AMINO ACID RESIDUES
;       TYPE:  AMINO ACID
;       STRANDEDNESS:  UNKNOWN
;       TOPOLOGY:  UNKNOWN
PCT-US93-08849A-1


  Query Match             1.6%;  Score 179.5;  DB 4;  Length 1693;
  Best Local Similarity   21.3%;  Pred. No. 4.4e-06;
  Matches  189;  Conservative   98;  Mismatches  320;  Indels  281;  Gaps   45;

Qy   1521 GKQLCMFRTC----SCGVQLD----VYNEA------TIATRFSNAFTFVDNLKGRSAVFF 1566
          |:||  :  |    | |  ||    |::|:      |   :  : |       |:    |
Db    425 GRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAIRKAVSKFCCFMKWLGQECTCF 484

Qy   1567 SKLGEGY---------TYNG-------------GSHVSSGWPRALEDILTAIKYPSVFD 1603
           :  ||           | |              ||:|  |   ||: :  |:  |:
Db    485 LQPAEGVVGDQGHDNEAYEGSDVDPAESAISDISGSYVVPG--TALQPLYQALDLPA--- 539

Qy   1604 HCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDY 1663
             : :  ::   |       :    | |    |:|   | |     | |:  |
Db    540 EIVARAGRLTATV-------KVSQVDGRIDCETLLGNKTFRTSF-VDGAVLETN------ 585

Qy   1664 FLMPCGFQRTHLH---SVNSIDEGRISLTFRATRR-------VFGVGRMLQLAGGVSDEK 1713
             | :| :|    | :::  |  |||: |:          .  |:      | |||
Db    586 -----GPERHNLSFDASQSTMAAGPFSLTYAASAAGLEVRYVAAGLDHRAVFAPGVSPRS 640

Qy   1714 SPGVP----------NQQPQ-----------SQGATRTITPKSGGKALSEGSGREVKGRS 1752
          :||            |:: |           :|      | | |       |     | |
Db    641 APGEVTAFCSALYRFNREAQRLSLTGNFWFHPEGLLGPFAPFSPGHVWE--SANPFCGES 698

Qy   1753 TY--SIWCEQDYVRKCEW------------LRADNPVMAL-----EPDYTPMTFEVVKTG 1793
          |     | | | |                  ||  |  |       || :| |      |
Db    699 TLYTRTWSEVDAVPSPAQPDLGFTSEPSIPSRAATPTPAAPLPPPAPDPSP-TLSAPARG 757

Qy   1794 TSEDAVVEYLKYLAIGIERTYRAL--------LMARNIAVTTAEGVLKVPN--------- 1836
                 :     |  | |       : | ::  :|   ::   |
Db    758 EPAPGATARAPAITHQTARHRRLLFTYPDGSKVFAGSLFESTCTWLVNASNVDHRPGGGL 817

Qy   1837 --QVYESLP------------GFHVYKSGTDLIFHSTQDGLRVRDLP------------- 1869
             |:  |              |   |       | |:      |:   |
Db    818 CHAFYQRYPASFDAASFVMRDGAAAYTLTPRPIIHAVAPDYRLEHNPKRLEAAYRETCSR 877

Qy   1870 -----YVLIAEKGIFTK--GKDVDAVV---ALGDNL-V--------------CDDILVFH 1905
               | |:   ||:    |   ||       || |::               |  : :
Db    878 LGTAAYPLLG-TGIYQVPIGPSFDAWERNHRPGDELYLPELAARWFEANRPTCPTLTITE 936

Qy   1906 D-------AINLIGALKVARC--------GMVGESFKSFEYKCYNAPPGGGKTTTLVDEF 1950
          |       || |  |  | |          |:|       :|:      || ||: ::
Db    937 DVARTANLAIELDSATDVGRACAGCRVTPGVV-------QYQFTAGVPGSGKSRSI---- 985

Qy   1951 VKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDE 2010
               : | :   |  : ::  | ::|     |  : |   :  |        :||::||
Db    986 -----TQADVDVVVVPTRELRNAWRRR-----GFAAFTPHTAARVT-----QGRRVVIDE 1030

Qy   2011 VHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPK-KESVVYTS-- 2067
               :  || | :    |:     || |||| |: |       |  ||  :  :  ||
Db   1031 APSLPPHLLLLHM--QRAATVHLLGDPNQIPAIDFEH------AGLVPAIRPDLAPTSWW 1082

Qy   2068 -KSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPIGTTDDVAEINA 2124
            ::||| ||| |     :||    ||    : :   |:|||   :  :|
Db   1083 HVTHRCPADVCEL-----IRG---AYPMIQTTSR---VLRSLFWGEPAVGQK-------- 1123

Qy   2125 DVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHI 2184
             |  ||  |:    |          :|||||| |:::  :  | |   : : : |
Db   1124 ---LVFTQAAKAANPGS----------VTVHEAQGATYTETTIIATADARGLIQSSRAHA 1170

Qy   2185 LVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAG 2232
          :| |:|||   |       | : ||          :|||::: |  ||
Db   1171 IVALTRHTEKCVIIDAPGLLRE-VG----------ISDAIVNNFFLAG 1207


RESULT   5
PCT-US93-08849-1
; Sequence 1, Application PC/TUS9308849
;  GENERAL INFORMATION:
;    APPLICANT:  Tsarev, Sergei A., Emerson,
;    APPLICANT:  Suzanne U., Purcell, Robert H.
;    TITLE OF INVENTION:  Recombinant Proteins Of
;    TITLE OF INVENTION:  A Pakistani Strain Of Hepatitis E And Their
;    TITLE OF INVENTION:  Use In Diagnostic Methods And Vaccines
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORGAN & FINNEGAN
;      STREET:  345 PARK AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10154
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/08849
;      FILING DATE:  17-SEP-1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/947,263
;      FILING DATE:  18-SEP-1992
;      NAME:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bork, Richard, W.
;      REGISTRATION NUMBER:  36,459
;      REFERENCE/DOCKET NUMBER:  2026-4032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 758-4800
;      TELEFAX:  (212) 751-6849
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 amino acid residues
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
PCT-US93-08849-1


  Query Match             1.6%;  Score 179.5;  DB 4;  Length 1693;
  Best Local Similarity   21.3%;  Pred. No. 4.4e-06;
  Matches  189;  Conservative   98;  Mismatches  320;  Indels  281;  Gaps   45;

Qy   1521 GKQLCMFRTC----SCGVQLD----VYNEA------TIATRFSNAFTFVDNLKGRSAVFF 1566
          |:||  :  |    | |  ||    |::|:      |   :  : |       |:    |
Db    425 GRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAIRKAVSKFCCFMKWLGQECTCF 484

Qy   1567 SKLGEGY---------TYNG-------------GSHVSSGWPRALEDILTAIKYPSVFD 1603
           :  ||           | |              ||:|  |   ||: :  |:  |:
```

```
Db    485 LQPAEGVVGDQGHDNEAYEGSDVDPAESAISDISGSYVVPG--TALQPLYQALDLPA--- 539

Qy   1604 HCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDY 1663
          : :  ::   |       :    | |    |:|   | |     | |: |
Db    540 EIVARAGRLTATV-------KVSQVDGRIDCETLLGNKTFRTSF-VDGAVLETN------ 585

Qy   1664 FLMPCGFQRTHLH---SVNSIDEGRISLTFRATRR-------VFGVGRMLQLAGGVSDEK 1713
               | :| :|    | :::  |  |||: |:           |:      | |||
Db    586 -----GPERHNLSFDASQSTMAAGPFSLTYAASAAGLEVRYVAAGLDHRAVFAPGVSPRS 640

Qy   1714 SPGVP----------NQQPQ-----------SQGATRTITPKSGGKALSEGSGREVKGRS 1752
          :||           |:: |           :|     | | |      |     | |
Db    641 APGEVTAFCSALYRFNREAQRLSLTGNFWFHPEGLLGPFAPFSPGHVWE--SANPFCGES 698

Qy   1753 TY--SIWCEQDYVRKCEW------------LRADNPVMAL-----EPDYTPMTFEVVKTG 1793
          |     | | | |                 ||  |  |       || :| | |      |
Db    699 TLYTRTWSEVDAVPSPAQPDLGFTSEPSIPSRAATPTPAAPLPPPAPDPSP-TLSAPARG 757

Qy   1794 TSEDAVVEYLKYLAIGIERTYRAL--------LMARNIAVTTAEGVLKVPN--------- 1836
                    :     | | |        : | ::  :|   ::   |
Db    758 EPAPGATARAPAITHQTARHRRLLFTYPDGSKVFAGSLFESTCTWLVNASNVDHRPGGGL 817

Qy   1837 --QVYESLP------------GFHVYKSGTDLIFHSTQDGLRVRDLP------------- 1869
             |:  |            |   |     ' | |:      |:   |
Db    818 CHAFYQRYPASFDAASFVMRDGAAAYTLTPRPIIHAVAPDYRLEHNPKRLEAAYRETCSR 877

Qy   1870 -----YVLIAEKGIFTK--GKDVDAVV---ALGDNLFV--------------CDDILVFH 1905
               | |:   ||:    |   ||        || |::               |  : :
Db    878 LGTAAYPLLG-TGIYQVPIGPSFDAWERNHRPGDELYLPELAARWFEANRPTCPTLTITE 936

Qy   1906 D-------AINLIGALKVARC--------GMVGESFKSFEYKCYNAPPGGGKTTTLVDEF 1950
          |       || |  |  | |          |:|        :|:      || ||: ::
Db    937 DVARTANLAIELDSATDVGRACAGCRVTPGVV-------QYQFTAGVPGSGKSRSI---- 985

Qy   1951 VKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDE 2010
               : | :   |  : ::  | ::|     |  : |   :  |        :||::||
Db    986 -----TQADVDVVVVPTRELRNAWRRR-----GFAAFTPHTAARVT-----QGRRVVIDE 1030

Qy   2011 VHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPK-KESVVYTS-- 2067
             :   || | :    |:     || |||| |: |       |  ||  :  :  ||
Db   1031 APSLPPHLLLLHM--QRAATVHLLGDPNQIPAIDFEH------AGLVPAIRPDLAPTSWW 1082

Qy   2068 -KSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPIGTTDDVAEINA 2124
            ::||| ||| |     :||    ||    : :   |:|||   :  :|
Db   1083 HVTHRCPADVCEL-----IRG---AYPMIQTTSR---VLRSLFWGEPAVGQK-------- 1123

Qy   2125 DVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHI 2184
             |  ||  |:    |          :|||||| |:::  :  |  |   : : : |
Db   1124 ---LVFTQAAKAANPGS----------VTVHEAQGATYTETTIIATADARGLIQSSRAHA 1170

Qy   2185 LVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAG 2232
          :| |:|||   |       | : ||          :|||::: |  ||
Db   1171 IVALTRHTEKCVIIDAPGLLRE-VG----------ISDAIVNNFFLAG 1207


RESULT   6
US-08-478-507-7
; Sequence 7, Application US/08478507
; Patent No. 6120988
;  GENERAL INFORMATION:
;    APPLICANT:  Reyes, Gregory R
;    APPLICANT:  Yarbough, Patrice O
;    APPLICANT:  Bradley, Daniel W
;    APPLICANT:  Krawczynski, Krzysztof Z
;    APPLICANT:  Tam, Albert
;    APPLICANT:  Fry, Kirk E
;    TITLE OF INVENTION:  DNA Sequences of Enterically Transmitted
;    TITLE OF INVENTION:  No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/478,507
;      FILING DATE:  07-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/279,823
;      FILING DATE:  25-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/681,078
;      FILING DATE:  05-APR-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/505,888
;      FILING DATE:  05-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/420,921
;      FILING DATE:  13-OCT-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/367,486
;      FILING DATE:  16-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/336,672
;      FILING DATE:  11-APR-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/208,997
;      FILING DATE:  17-JUN-1988
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0183.22
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 324-0880
;      TELEFAX:  (650) 324-0960
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1693 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-478-507-7


  Query Match             1.5%;  Score 169.5;  DB 3;  Length 1693;
  Best Local Similarity  20.1%;  Pred. No. 3.5e-05;
  Matches  179;  Conservative  103;  Mismatches  323;  Indels  285;  Gaps   41;

Qy   1521 GKQLCMFRTC----SCGVQLD----VYNEA------TIATRFSNAFTFVDNLKGRSAVFF 1566
          |:||  :  |    | |  ||    |::|:      |   :  : |       |:    |
Db    425 GRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAIRKALSKFCCFMKWLGQECTCF 484

Qy   1567 SKLGEG---------YTYNG-------------GSHVSSGWPRALEDILTAIKYPSVFD 1603
             :  ||           | |              ||:|  |   ||: :  |:  |:
Db    485 LQPAEGAVGDQGHDNEAYEGSDVDPAESAISDISGSYVVPG--TALQPLYQALDLPA--- 539

Qy   1604 HCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDY 1663
             : :  ::   |       :    | |    |:|   | |     | |: |
Db    540 EIVARAGRLTATV-------KVSQVDGRIDCETLLGNKTFRTSF-VDGAVLETN------ 585

Qy   1664 FLMPCGFQRTHLH---SVNSIDEGRISLTFRATRR-------VFGVGRMLQLAGGVSDEK 1713
               | :| :|    | :::  |  |||: |:           |:      | |||
Db    586 -----GPERHNLSFDASQSTMAAGPFSLTYAASAAGLEVRYVAAGLDHRAVFAPGVSPRS 640

Qy   1714 SPG--------------------VPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRS 1752
          :||                     : |    :|      | | |      |     | |
Db    641 APGEVTAFCSALYRFNREAQRHSLIGNLWFHPEGLIGLFAPFSPGHVWE--SANPFCGES 698

Qy   1753 TY--SIWCEQDYVRKCEW------------LRADNPVMAL-----EPDYTPMTFEVVKTG 1793
```

```
Db    699 TLYTRTWSEVDAVSSPARPDLGFMSEPSIPSRAATPTLAAPLPPPAPDPSP--------P 750

Qy   1794 TSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESL---------- 1842

Db    751 PSAPALAEPASGATAGAPAITHQTARHRRLLFTYPDGSKVFAGSLFESTCTWLVNASNVD 810

Qy   1843 --PG-------FHVYKSGTDLIFHSTQDG-----LRVRDLPYVLIAEKGIFTKGKDVDAV 1888

Db    811 HRPGGGLCHAFYQRYPASFDAASFVMRDGAAAYTLTPRPIIHAVAPDYRLEHNPKRLEA- 869

Qy   1889 VALGDNLFVCDDILVFHDAINLIGALKVARCGM------VGESFKSFEYKCYNAPPGG-- 1940

Db    870 -------------AYRETCSRLGTAAYPLLGTGIYQVPIGPSFDAWE---RNHRPGDEL 912

Qy   1941 ----------------GKTTTLVDEFVKSPNSTATITANVGSSEDINMAVK--------- 1975

Db    913 YLPELAARWFEANRPTRPTLTITEDVAR----TANLAIELDSATDVGRACAGCRVTPGVV 968

Qy   1976 --KRDPNLEGLNSATTVNSRVVNFIV------------RGMY-------------KRVLV 2008

Db    969 QYQFTAGVPGSGKSRSITQADVDVVVVPTRELRNAWRRRGFAAFTPHTAARVTQGRRVVI 1028

Qy   2009 DEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPK-KESVVYTS 2067

Db   1029 DEAPSLPPHLLLLHM--QRAATVHLLGDPNQIPAIDFEH------AGLVPAIRPDLGPTS 1080

Qy   2068 ---KSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSL--SKRPIGTTDDVAEI 2122

Db   1081 WWHVTHRWPADVCEL-----IRG---AYPMIQTTSR---VLRSLFWGEPAVGQK------ 1123

Qy   2123 NADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQP 2182

Db   1124 -----LVFTQAAKPANPGS----------VTVHEAQGATYTETTIIATADARGLIQSSRA 1168

Qy   2183 HILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAG 2232

Db   1169 HAIVALTRHTEKCVIIDAPGLLRE-VG----------ISDAIVNNFFLAG 1207


RESULT   7
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-15


  Query Match             1.3%;  Score 150.5;  DB 2;  Length 1222;
  Best Local Similarity   20.0%;  Pred. No. 0.00095;
  Matches  185;  Conservative  110;  Mismatches  345;  Indels  287;  Gaps   45;

Qy   1262 DETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNG-EVT 1320

Db    164 DFALVFKRTVDKVEGETP----EEAAFVKAINNTTVEVTFEEEVTNVQALNFKIEGLEIK 219

Qy   1321 GEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSG--SGTADAM-AFVESVKKGVDDVFH 1377

Db    220 NASVKQTNKKVVVLTTEAQTADKEYVLTLDGETIGGFKGVAAVVPTKVELVSSAV----- 274

Qy   1378 QQSSGETAREVEVDGK-----------------------GLLPESVVGEAPTQERG--- 1410

Db    275 ---QGKLGQEVKVQAKVTVAEGQSKAGIPVTFTVPGNNNDGVVP-TLTGEALTNEEGIAT 330

Qy   1411 ----RAADGNTAQTAVNEGDR------------------EPVQSSLVSSPQADIPKVTQ 1447

Db    331 YSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGVDTILSVEEVTTGASVNNGANKTYKVTY 390

Qy   1448 SEVHAQK-EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVE 1506

Db    391 KNPKTGKPEANKTFNVGFVENMN-VTSDKVA-NATVNGVKALQLSNGTA--LDAAQITTD 446

Qy   1507 QPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFF 1566

Db    447 SKGEATFTVSGTNAAVTPVVY-------DLHSTNNSTSNKKYSASALQTT----ASKVTF 495

Qy   1567 SKLGEGYTY------NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGG------ 1614

Db    496 AALQAEYTIELTRADNAGE------------VAAIGATNGREYKVIVKDKAGNLAKNEI 542

Qy   1615 -GVPFHADDEEC--------YPSDNPILTVNLVG-KA-NFSTKCR-KCGKVMVINVASGD 1662

Db    543 VNVAFNEDKDRVISTVTNAKFVDTDPDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVN 602

Qy   1663 YFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPN--- 1719

Db    603 DYATPIAW-----IDINTSDAKQ----------------------GDLDEGEPKAVAPIS 635

Qy   1720 --QQPQSQGA----------TRTITPKSGGKA------LSEGSGREVKG----RSTYSIW 1757

Db    636 YFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAELVNQSGKKVTGTSIKKATYTI- 694

Qy   1758 CEQDYVRKCEWLRADNPVMALEPDYTPMTFEVV--KTGTSEDAVVEYLKYLAIGIERTYR 1815

Db    695 ----YNTGANDIKVDNQVISPNRSYT-VTYEATLSSTGT---------------------- 728

Qy   1816 ALLMARNIAVTTAEG---VLKVPNQVYESLPGFHVYKSGTDLIF---HSTQDGLRVRDLP 1869

Db    729 VITPAKNLEVTSVDGKTTAVKV------IATGIAVNTDGKDYAFTAKEATATFTATNEVP 782

Qy   1870 YVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSF 1929

Db    783 N---SYTGVATQFNTADS----GSN----------SNSIWFAGKNPVKYAGVSGKTYKYF 825

Qy   1930 -------------EYKCYNAPPGGGKTT--------TLVDEFVKSPNSTA---------- 1958

Db    826 GANGNEVFGEAAWEALLTQYATEGQKVTISYNVDGDTVTFKVISAVNSSTEAIKPVAPTT 885

Qy   1959 ---------TITANVGSSEDINMAVKKRDPNL--EGLN-SATTVNSRVVNFIVRGMYKRV 2006

Db    886 PAAPTTGALTLTPAAGGLVDLTTATNTLGISLADADLNVSATTVDTATVS---------- 935

Qy   2007 LVDEVHMMHQGLLQLGVFATGASEGLF 2033

Db    936 LKDSAN----NSLSLTLVETGANTGVF 958


RESULT   8
US-08-682-517-9
; Sequence 9, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-9


  Query Match             1.3%;  Score 150.5;  DB 2;  Length 1252;
  Best Local Similarity   20.0%;  Pred. No. 0.00099;
  Matches  185;  Conservative  110;  Mismatches  345;  Indels  287;  Gaps   45;

Qy   1262 DETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNG-EVT 1320
          |  | : |  |  ||  |   ||      : :|  |:  : ||||: :     | |:
Db    194 DFALVFKRTVDKVEGETP----EEAAFVKAINNTTVEVTFEEEVTNVQALNFKIEGLEIK 249

Qy   1321 GEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSG--SGTADAM-AFVESVKKGVDDVFH 1377
             : : |     | : |    |    |:  |   | |  :   || |   |
Db    250 NASVKQTNKKVVVLTTEAQTADKEYVLTLDGETIGGFKGVAAVVPTKVELVSSAV----- 304

Qy   1378 QQSSGETAREVEVDGK-----------------------GLLPESVVGEAPTQERG--- 1410
             |:  :||:|  |                         |::| :: ||| | | |
Db    305 ---QGKLGQEVKVQAKVTVAEGQSKAGIPVTFTVPGNNNDGVVP-TLTGEALTNEEGIAT 360

Qy   1411 ----RAADGNTAQTAVNEGDR------------------EPVQSSLVSSPQADIPKVTQ 1447
             |  :|   ||  |||                    |   : |::     |||
Db    361 YSYTRYKEGTDEVTAYATGDRSKFSLGYVFWGVDTILSVEEVTTGASVNNGANKTYKVTY 420

Qy   1448 SEVHAQK-EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVE 1506
             | |  :   :  |     :  :| | :|  ||| :      |   :  |:  :
Db    421 KNPKTGKPEANKTFNVGFVENMN-VTSDKVA-NATVNGVKALQLSNGTA--LDAAQITTD 476

Qy   1507 QPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFF 1566
             : : |::   |     ::        |   | :|   ::| :          | | |
Db    477 SKGEATFTVSGTNAAVTPVVY-------DLHSTNNSTSNKKYSASALQTT----ASKVTF 525

Qy   1567 SKLGEGYTY------NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGG------ 1614
          : |   ||       | |          : ||   :  :: :: | | |
Db    526 AALQAEYTIELTRADNAGE-------------VAAIGATNGREYKVIVKDKAGNLAKNEI 572

Qy   1615 -GVPFHADDEEC--------YPSDNPILTVNLVG-KA-NFSTKCR-KGGKVMVINVASGD 1662
            | |: | :          :   :|   |  | ||   | |   ||    ||   : :
Db    573 VNVAFNEDKDRVISTVTNAKFVDTDPDTAVYFTGDKAKQISVKTNDKGEATFVIGSDTVN 632

Qy   1663 YFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPN--- 1719
            :  |  :       :|: |  :                       |  ||  |
Db    633 DYATPIAW-----IDINTSDAKQ----------------------GDLDEGEPKAVAPIS 665

Qy   1720 --QQPQSQGA----------TRTITPKSGGKA------LSEGSGREVKG----RSTYSIW 1757
            | |  |:          : :|   | :       |   ||::| |    ::||:|
Db    666 YFQAPYLDGSAIKAYKKSDLNKAVTKFDGSETAVFAAELVNQSGKKVTGTSIKKATYTI- 724

Qy   1758 CEQDYVRKCEWLRADNPVMALEPDYTPMTFEVV--KTGTSEDAVVEYLKYLAIGIERTYR 1815
             |      :: || |::    || :|:|'     |||
Db    725 ----YNTGANDIKVDNQVISPNRSYT-VTYEATLSSTGT---------------------- 758

Qy   1816 ALLMARNIAVTTAEG---VLKVPNQVYESLPGFHVYKSGTDLIF---HSTQDGLRVRDLP 1869
            :  |:|: ||: :|    :||         |  |   | |  |    :|       ::|
Db    759 VITPAKNLEVTSVDGKTTAVKV------IATGIAVNTDGKDYAFTAKEATATFTATNEVP 812

Qy   1870 YVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSF 1929
            :  |: |:    |:    | |           ::|   |   |   |: |:::| |
Db    813 N---SYTGVATQFNTADS----GSN----------SNSIWFAGKNPVKYAGVSGKTYKYF 855

Qy   1930 -------------EYKCYNAPPGGGKTT--------TLVDEFVKSPNSTA---------- 1958
                       |         | | |        |:  : : : ||:
```

```
Db    856 GANGNEVFGEAAWEALLTQYATEGQKVTISYNVDGDTVTFKVISAVNSSTEAIKPVAPTT 915

Qy   1959 ---------TITANVGSSEDINMAVKKRDPNL--EGLN-SATTVNSRVVNFIVRGMYKRV 2006
                    |:|   |   |:  |      :|    || |||||::  |:
Db    916 PAAPTTGALTLTPAAGGLVDLTTATNTLGISLADADLNVSATTVDTATVS---------- 965

Qy   2007 LVDEVHMMHQGLLQLGVFATGASEGLF 2033
          | |  :      | | :  |||: |:|
Db    966 LKDSAN----NSLSLTLVETGANTGVF 988


RESULT   9
5183745-6
;Patent No. 5183745
;    APPLICANT: DANCHIN, ANTOINE;GLASER, PHILLIPPE;KRIN, EVELYN;
;BARZU, OCTAVIEN;LADANT, DANIEL;ULLMAN, AGNES
;    TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
;BIOLOGICAL USES
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,541
;      FILING DATE: 25-OCT-1989
;SEQ ID NO:6:
;      LENGTH: 1794
5183745-6


  Query Match             1.3%;  Score 149.5;  DB 5;  Length 1794;
  Best Local Similarity   20.0%;  Pred. No. 0.0025;
  Matches  262;  Conservative  162;  Mismatches  506;  Indels  381;  Gaps   61;

Qy    522 KAGHVNC-----HVCNPVLD-VKDVKRRINEILFLSTAGGDSYVSSDLLTE-AASKSVSY 574
          :||: |       :   ||| :| | :  |  |           |: |: |  |:| :
Db      6 QAGYANAADRESGIPAAVLDGIKAVAKEKNATLMFRLVNPH---STSLIAEGVATKGLGV 62

Qy    575 CSRESQNCDSRADAGF------MVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGN 628
           :: |   |    ||:      :  ::  :|: :| | :   :            | :|:
Db     63 HAKSS---DWGLQAGYIPVNPNLSKLFGRAPEVIARADNDVNS----------SLAHGH 108

Qy    629 GEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSFSNVSGFFTFSY-VRTSSGNVFKLEYE 687
           | |   || |   |||        |        || :|:  | : |: :|    : ::|
Db    109 TAVDL----TLSKERLDYLRQAGLVTGMADGVVASNHAGYEQFEFRVKETSDGRYAVQYR 164

Qy    688 ------------GYRCG-----------YHHLTMCRAQKSPGTEVTYRSLVPSFVGKS 722
                       |   |                ||:  |   |  : ||    |  :: ::
Db    165 RKGGDDFEAVKVIGNAAGIPLTADIDMFAIMPHLSNFR--DSARSSVTSGDSVTDYLART 222

Qy    723 LVFIPVVAGSSVSFKTIVLDSDFVDRIYSY----ALNTIGTFENRTFEY----AVGAVR- 773
                  : |  |  || : :| :        | : :||   | | |     :| :
Db    223 --------RRAASEATGGLDRERIDLLQKIARAGARSAVGTEARRQFRYDGDMNIGVITD 274

Qy    774 ---------SQKTHVITGSRVV-HSKVDISP----DDMWGLVVAVMAQAIKDRAKSIRSY 819
                    ::: | :    || |     :|    |:   :| |     :  | :
Db    275 FELEVRNALNRRAHAVGAQDVVQHGTEQNNPFPEADEKIFVVSATGESQMLTRGQ----- 329

Qy    820 NFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAF----IR 875
            :|   |   |   :|::      : |  |  |::  |       : |    |    :
Db    330 --LKEYIGQQRGEGYVFYE------NRAYGVAGKSLFDDGLGAAPGVPSGRSKFSPDVLE 381

Qy    876 KVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMK 935
          || |  :   : || :| :    :|       |     | ||   | |       |
Db    382 TVPASPGLRRPSLGAVERQD--SGYD--------SLDGVGSRSFSLGEVSD-------MA 424

Qy    936 TEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATK---A 992
             :  | :|   ::|    :  :   : |   | :|       :|  | ||  :   |
Db    425 AVEAAELEMTRQVLHAGARQDDAEPGVSGASAHWGQRAL------QGAQAVAAAQRLVHA 478

Qy    993 FSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFW 1052
           : :| |   ||                        |:|         ::|  | |  |
Db    479 IALMTQFGRAGS--------------------TNTP-----QEAASLSAAVFGLGEAS 511

Qy   1053 SEVSLNDFTTFVLRNKVLIGIFVAS------LGAAPIAWKYRRGIAANARRYAGSSYETL 1106
          | |:           : : | |  |       | |  |     ||||     || |
```

```
Db    512 SAVA-----------ETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVG--AGMSLTDD 558

Qy   1107 SSLSSQAAGGLR---GLTSSTVSGGS-----LVVRRGFSSAVTVTRATVAKRQVPLA--- 1155
          :   :|| |      ||  ||   |     |   || :| : |  |:          ||
Db    559 APAGQKAAAGAEIALQLTGGTVELASSIALALAAARGVTSGLQVAGASAGAAAGALAAAL 618

Qy   1156 --------------------LLSFSTSYAISGCSMLG-IWAHALPRHLMFFFGLGTLLGA 1194
                               |   |::|   | ::|  ::              | : |
Db    619 SPMEIYGLVQQSHYADQLDKLAQESSAYGYEGDALLAQLYRDKTAAE-------GAVAGV 671

Qy   1195 RASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPP 1254
           |  :|    | : :  |   ||  |  |: :   :|  :: |:||:    | :||
Db    672 SAVLST---VGAAVSIAAAASVV--GAPVAVVTSLLTGALNGILRGVQQPIIEKLA---- 722

Qy   1255 VEGRNVYDETLRYYRDFDYDEGAGPSG-------TQHEAVPGDD-------------NDG 1294
           | |   :         ||  ||           :|| :   |                 |
Db    723 ----NDYARKI--------DELGGPQAYFEKNLQARHEQLANSDGLRKMLADLQAGWNAS 770

Qy   1295 STSSVSSYDV-------------VTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVA 1341
          |   | : ::               |:: | :  :  | ||      |       | :  |
Db    771 SVIGVQTTEISKSALELAAITGNADNLKSVDVFVDRFVQGERVAGQP------VVLDVAA 824

Qy   1342 PSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSG--ETAREVEVDGKGLLPES 1399
           :| |:|:        |: |:  :    ::   :  :|  |    ||: ||
Db    825 GGIDIASRKGE------RPALTFITPLAAPGEEQRRRTKTGKSEFTTFVEIVGK------ 872

Qy   1400 VVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQE 1459
               |:| |  ||  | | :                 |: ||    | |   :|
Db    873 -------QDRWRIRDG-AADTTI-----------------DLAKVVSQLVDANGVLKHS 906

Qy   1460 VPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKA----VAHVAEKKQVQVEQPKQRSLTI 1515
          : |  : |    |        |    :|   | |    |:: |          :| |:|:
Db    907 IKLDVIGGDGDDV-------VLANASRIHYDGGAGTNTVSYAA--------LGRQDSITV 951

Qy   1516 N-EGK---AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGE 1571
          : :|:     |||             :|| |  :||: :      :|:: |  |  :::|:
Db    952 SADGERFNVRKQL---------NNANVYREG-VATQTTAYGKRTENVQYRH-VELARVGQ 1000

Qy   1572 GYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNP 1631
           :  ||      |  | :|     :  |: |      ||     ||
Db   1001 VVEVDTLEHVQHIIGGAGNDSITG----NAHDNFLA-----GGS----GDD--------- 1038

Qy   1632 ILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSID 1682
               | |  |     |:  ||  |  | ||    |      |     :|
Db   1039 ----RLDGGAGNDTLVGGEGQNTVIGGAGDDVFLQDLGVWSNQLDGGAGVD 1085


RESULT 10
5183745-2
;Patent No. 5183745
;    APPLICANT: DANCHIN, ANTOINE;GLASER, PHILLIPPE;KRIN, EVELYN;
;BARZU, OCTAVIEN;LADANT, DANIEL;ULLMAN, AGNES
;    TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
;BIOLOGICAL USES
;    NUMBER OF SEQUENCES: 13
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/426,541
;      FILING DATE: 25-OCT-1989
;SEQ ID NO:2:
;      LENGTH: 1489
5183745-2


  Query Match             1.2%;  Score 142.5;  DB 5;  Length 1489;
  Best Local Similarity   19.8%;  Pred. No. 0.0073;
  Matches  260;  Conservative  163;  Mismatches  507;  Indels  381;  Gaps   61;

Qy    522 KAGHVNC-----HVCNPVLD-VKDVKRRINEILFLSTAGGDSYVSSDLLTE-AASKSVSY 574
          :||: |       :   |||  :| | :  |  |           |: |: |  |:| :
Db      7 QAGYANAADRESGIPAAVLDGIKAVAKEKNATLMFRLVNPH---STSLIAEGVATKGLGV 63

Qy    575 CSRESQNCDSRADAGF------MVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGN 628
           :: |   |    ||:      :  ::  :|: :| | :   :            | :|:

Db     64 HAKSS---DWGLQAGYIPVNPNLSKLFGRAPEVIARADNDVNS----------SLAHGH 109

Qy    629 GEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSFSNVSGFFTFSY-VRTSSGNVFKLEYE 687
            | |    || |   |||        |       || :|:  | : |: :|   : ::|
Db    110 TAVDL----TLSKERLDYLRQAGLVTGMADGVVASNHAGYEQFEFRVKETSDGRYAVQYR 165

Qy    688 -------------GYRCG------------YHHLTMCRAQKSPGTEVTYRSLVPSFVGKS 722
                        |   |              ||:  |   |  : ||    |  :: ::
Db    166 RKGGDDFEAVKVIGNAAGIPLTADIDMFAIMPHLSNFR--DSARSSVTSGDSVTDYLART 223

Qy    723 LVFIPVVAGSSVSFKTIVLDSDFVDRIYSY----ALNTIGTFENRTFEY----AVGAVR- 773
                  : |  |  || : :| ::       | : :||   | | |     :| :
Db    224 --------RRAASEATGGLDRERIDLLWKIARAGARSAVGTEARRQFRYDGDMNIGVITD 275

Qy    774 ---------SQKTHVITGSRVV-HSKVDISP----DDMWGLVVAVMAQAIKDRAKSIRSY 819
                  ::: | :    || |     :|    |:   :| |     :  | :
Db    276 FELEVRNALNRRAHAVGAQDVVQHGTEQNNPFPEADEKIFVVSATGESGMLTRGQ----- 330

Qy    820 NFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAF----IR 875
            :    |   |  :|::       :    |  |::  |       : |    |    :
Db    331 --LDEYIGQQRGEGYVFYE------NRQYGVAGKSLFDDGLGAAPGVPSGRSKFSPDVLE 382

Qy    876 KVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMK 935
           || |  :   : || :| :    :|          |      | ||   | |       |
Db    383 TVPASPGLRRPSLGAVERQD--SGYD--------SLDGVGSRSFSLGEVSD-------MA 425

Qy    936 TEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATK---A 992
            :   | :|   ::|    : :    : |    | :|       :|  |  || :   |
Db    426 AVEAAELEMTRQVLHAGARQDDAEPGVSGASAHWGQRAL------QGAQAVAAAQRLVHA 479

Qy    993 FSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFW 1052
            : :| |   ||                        |:|         ::|  | |  |
Db    480 IALMTQFGRAGS----------------------TNTP-----QEAASLSAAVFGLGEAS 512

Qy   1053 SEVSLNDFTTFVLRNKVLIGIFVAS------LGAAPIAWKYRRGIAANARRYAGSSYETL 1106
           | |:           : : | |  |       | |  |     ||||     || |
Db    513 SAVA-----------ETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVG--AGMSLTDD 559

Qy   1107 SSLSSQAAGGLR---GLTSSTVSGGS-----LVVRRGFSSAVTVTRATVAKRQVPLA--- 1155
          :    :|| |      ||  ||   |     |   || :| : |  |:        ||
Db    560 APAGQKAAAGAEIALQLTGGTVELASSIALALAAARGVTSGLQVAGASAGAAAGALAAAL 619

Qy   1156 --------------------LLSFSTSYAISGCSMLG-IWAHALPRHLMFFFGLGTLLGA 1194
                               |   |::|   | ::|  ::              | : |
Db    620 SPMEIYGLVQQSHYADQLDKLAQESSAYGYEGDALLAQLYRDKTAAE-------GAVAGV 672

Qy   1195 RASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPP 1254
           |  :|    | : :  |   ||  |  |: :   :|  :: |:||:    | :||
Db    673 SAVLST---VGAAVSIAAAASVV--GAPVAVVTSLLTGALNGILRGVQQPIIEKLA---- 723

Qy   1255 VEGRNVYDETLRYYRDFDYDEGAGPSG-------TQHEAVPGDD-------------NDG 1294
           | |   :         ||  ||           :|| :   |                 |
Db    724 ----NDYARKI--------DELGGPQAYFEKNLQARHEQLANSDGLRKMLADLQAGWNAS 771

Qy   1295 STSSVSSYDV-------------VTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVA 1341
          |   | : ::               |:: | :  :  | ||      |       | :  |
Db    772 SVIGVQTTEISKSALELAAITGNADNLKSVDVFVDRFVQGERVAGQP------VVLDVAA 825

Qy   1342 PSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSG--ETAREVEVDGKGLLPES 1399
           :| |:|:        |: |:  :    ::   :  :|  |    ||: ||
Db    826 GGIDIASRKGE------RPALTFITPLAAPGEEQRRRTKTGKSEFTTFVEIVGK------ 873

Qy   1400 VVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQE 1459
               |:| |  ||  | | :                 |: ||    | |   :|
Db    874 -------QDRWRIRDG-AADTTI-----------------DLAKVVSQLVDANGVLKHS 907

Qy   1460 VPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKA----VAHVAEKKQVQVEQPKQRSLTI 1515
          : |  : |    |        |    :|   | |    |:: |          :| |:|:
Db    908 IKLDVIGGDGDDV-------VLANASRIHYDGGAGTNTVSYAA--------LGRQDSITV 952

Qy   1516 N-EGK---AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGE 1571
          : :|:     |||             :|| |  :||: :      :|:: |  |  :::|:
Db    953 SADGERFNVRKQL---------NNANVYREG-VATQTTAYGKRTENVQYRH-VELARVGQ 1001
```

```
Qy   1572 GYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNP 1631
Db   1002 VVEVDTLEHVQHIIGGAGNDSITG----NAHDNFLA-----GGS----GDD--------- 1039

Qy   1632 ILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSID 1682
Db   1040 ----RLDGGAGNDTLVGGEGQNTVIGGAGDDVFLQDLGVWSNQLDGGAGVD 1086


RESULT 11
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1848 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-6


  Query Match             1.2%;  Score 142.5;  DB 4;  Length 1848;
  Best Local Similarity   18.3%;  Pred. No. 0.011;
  Matches  250;  Conservative  151;  Mismatches  435;  Indels  531;  Gaps   63;

Qy    532 NP--VLDVKDVKRRINEILFLSTAG----GDSYVSSDLLTEA----------ASKSVSY 574
Db     91 NPQYVVGVKHVSNGVSELHFGNLNGNMNNGNAKSHRDVSSEENRYYTVEKNNFPTENVTS 150

Qy    575 CSRESQNCDSRADAGFM--VD--VYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNG- 629
Db    151 FTKEEQDAQKRREDYYMPRLDKFVTEVAPIEASTANNNKGE--YNNSDKYPAFVRLGSGT 208

Qy    630 ----------EVYLEELDTLVKREGDYL--------------AYNVGQCGEMY------ 658
Db    209 QFIYKKGSRYQLILTEKD----KQGNLLRNWDVGGDNLELVGNAYTYGIAGTPYKVNHEN 264

Qy    659 ----------EHS----------FSNV-----SGFFTFSYVRTSSGNVFKLEYEGYRCG 692
Db    265 NGLIGFGNSKEEHSDPKGILSQDPLTNYAVLGDSGSPLFVYDREKGKWLFLGSYD-FWAG 323

Qy    693 YHHLTM--CRAQKSPGTEVTYRSL-VPSFVGKSLVFIPVVAGSSVSFKTIV-----LDSD 744
Db    324 YNKKSWQEWNIYKHEFAEKIYQQYSAGSLIGSNTQYTWQATGST---STITGGGEPLSVD 380

Qy    745 FVD------RIYSYALNTIGTFE-NRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDM 797
Db    381 LTDGKDKPNHGKSITLKGSGTLTLNNHIDQGAGGLFFEGDYEVKGT---------SDSTT 431

Qy    798 W---GLVVA-------VMAQAIKDRAKSIRSYNFIKASEGSLAGVFKL---------- 835
Db    432 WKGAGVSVADGKTVTWKVHNPKYDRLAKIGKGTLVVEGKGKNEGLLKVGDGTVILKQKAD 491

Qy    836 ------FFQTVG------------DCFSNAVSVY------------------------ 851
Db    492 ANNKVQAFSQVGIVSGRSTLVLNDDKQVDPNSIYFGFRGGRLDLNGNSLTFDHIRNIDDG 551

Qy    852 AKAMVHDNFNVL------ETLMSMPRAF--------------IRKVP---------GSV 881
Db    552 ARVVNHNMTNTSNITITGESLITNPNTITSYNIEAQDDDHPLRIRSIPYRQLYFNQDNRS 611

Qy    882 VVTICTSGASDRLEL--------------RGAFDISKETFGRKLKNSRLRVFSRAIVEDS 927
Db    612 YYTL-KKGASTRSELPQNSGESNENWLYMGRTSDAAKRNVMNHINNERMNGFNGYFGEEE 670

Qy    928 IKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKG-LVAR 986
Db    671 TKA-----TQNGK-LNVTFN----------GKSDQNRFLLTGGTNLNGDLNVEKGTLFLS 714

Qy    987 GAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVA 1046
Db    715 GRPTPHARDIAGISST---------KKDPHF-----TEN---------------------- 739

Qy   1047 GTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETL 1106
Db    740 ------NEVVVED--DWINRN------FKAT----------TMNVTGNASLYSGRNVANI 775

Qy   1107 SSLSSQAAGGLRGLTSST--------VSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLS 1158
Db    776 TS----------NITASNNAQVHIGYKTGDTVCVRSDYTGYVTCHNSNLSEK----ALNS 821

Qy   1159 FSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANT----------WKFGGFSN 1208
Db    822 FNPTNLRGNVN--------LTENASFTLGKANLFGTIQSIGTSQVNLKENSHWHLTGNSN 873

Qy   1209 -------------------------NWCAVPEVVWRG------------------KSVS 1224
Db    874 VNQLNLTNGHIHLNAQNDANKVTTYNTLTVNSLSGNGSFYYWVDFTNNKSNKVVVNKSAT 933

Qy   1225 SLL---------------------------LPITLGVSLIIRG-----LLNDTIPQLAY 1251
Db    934 GNFTLQVADKTGEPNHNELTLFDASNATRNNLEVTLANGSVDRGAWKYKLRNVNGRYDLY 993

Qy   1252 VPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDV 1311
Db    994 NPEVEKRN---------QTVDTTNITTPNDIQADAPSAQSNNEEIARVE-----TPVPPP 1039

Qy   1312 GISTNGEVTGEEETHSPRSVQYTYVEE-------EVAPSAAVAERQGDPSG----SGTAD 1360
Db   1040 APATESAIASEQPETRPAETAQPAMEETNTANSTETAPKSDTATQTENPNSESVPSETTE 1099

Qy   1361 AMA-----FVESVKKGVDDVFH-QQSSGETAREVEVDGKGLLPESVVGEAPTQ-ERGRAA 1413
Db   1100 KVAENPPQENETVAKNEQEATEPTPQNGEVAKE----------DQPTVEANTQTNEATQS 1149

Qy   1414 DGNTAQTAVNEGDREPVQSSLVS--SPQADIPKVTQSEVHAQKEVK-------------- 1457
Db   1150 EGKTEETQTAETKSEPTESVTVSENQPEKTVSQSTEDKVVVEKEEKAKVETEETQKAPQV 1209

Qy   1458 ------------QEVPL---------------ATVSGA---------------------- 1468
Db   1210 TSKEPPKQAEPAPEEVPTDTNAEEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSEKTN 1269

Qy   1469 ----TPIVDEKPA--PSVTTRGVKI-IDKGKAVAHVAEKKQVQVEQP 1508
```

```
                ||:| |  |  |: |    |:  :| : |  || ::  : |||
Db   1270 AEPVTPVVSENTATQPTETEETAKVEKEKTQEVPQVASQESPKQEQP 1316


RESULT  12
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen I.
;    APPLICANT:  St. Geme III, Joseph W.
;    TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/409,995
;      FILING DATE:  24-MAR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
US-08-409-995-4


  Query Match             1.2%;  Score 138.5;  DB 1;  Length 1912;
  Best Local Similarity   18.1%;  Pred. No. 0.027;
  Matches  315;  Conservative  208;  Mismatches  598;  Indels  617;  Gaps   77;

Qy     627 GNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSFS--------------------NVS 666
           ||| |:|  ||:       |   |   | :   ||:                     |:
Db     168 GNGNVHLNGLDST-------LPDAVTNTGVLSSSSFTPNDVEKTRAATVKDVLNAGWNIK 220

Qy     667 GFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFI 726
           |       :|: |||  ::
Db     221 G------AKTAGGNVESVDL---------------------------------------- 234

Qy     727 PVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITG---- 782
           | | ::| |  |  | : :| : :   |   |     |      :: :   : ||
Db     235 -VSAYNNVEF--ITGDKNTLDVVLTAKENXKTT--EVKFTPKTSVIKEKDGKLFTGKENN 289

Qy     783 --SRVVHSKVDISPDDMWGLVVAVMAQAIKD-------RAKSIRSYN------------- 820
             ::|  :    : |:  |||    |:|: |       | |:  :
Db     290 DTNKVTSNTATDNTDEGNGLVT---AKAVIDAVNKAGWRVKTTTANGQNGDFATVASGTN 346

Qy     821 -FIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPG 879
            :: :|: | | |       |   | ::|   | | |                 :|:
Db     347 VTFESGDGTTASVTK-------DTNGNGITVKYDAKVGDGLKFDSD---------KKIVA 390

Qy     880 SVVVTICTSGASDRLELRGAFDISKETFGRKLKNS---------------------RLR 917
                 | |           :|:||   :|| |:                        |
Db     391 DTTALTVTGGK--------VAEIAKEDDKKKLVNAGDLVTALGNLSWKAKAEADTDGALE 442

Qy     918 VFSRAIVEDSIKVMKAMKTEDGKPLPITED------------SVYAFIMGNVSN----- 959
             |:   :  :|  : :  : || | : :|              : :  :|  :|
Db     443 GISK---DQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSITLGGTTNGGNDA 499

Qy     960 --------VHCTRAGLLGGSKATVVSSVSKGLVARG--AAIKAFSGITSFFSTGSLFYDR 1009
                  :  | ||  ||:  |   ||:|  :  |  | |   ||: ::
Db     500 KTVINKDGLTITPAG-NGGTTGTNTISVTKDGIKAGNKAITNVASGLRAY---------- 548

Qy    1010 GLTEDERLDALVRTENAINSPV-----GILETSRVAVSK--------------------V 1044
              :|   | |  :   :|  |     |:|  :    :|                      |
Db     549 ---DDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLRKLGWV 605

Qy    1045 VA---GTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGS 1101
             |:   ||||  ::|   |   |               |||  :  |             |
Db     606 VSTKNGTKEESNQVKQADEVLF------------TGAGAATVTSKSEN----------GK 643

Qy    1102 SYETLSSLSSQAAGGLRG-----------------LT-----SSTVSGGSLVVRRGFSSA 1139
             |:|   ::|  ||                    ||     ::   ||   |: | : |
Db     644 HTITVSVAETKADCGLEKDGDTIKLKVDNQNTDNVLTVGNNGTAVTKGGFETVKTGATDA 703

Qy    1140 ----VTVTRAT--------VAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFG 1187
                |||  ||           : |  |: | :|
Db     704 DRGKVTVKDATANDADKKVATVKDVATAINSAATFVKTE--------------------N 743

Qy    1188 LGTLL-----------GARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSL 1236
             | | :            |  : :|  |     |      :|   ||:::       |  :|
Db     744 LTTSIDEDNPTDNGKDDALKAGDTLTFKAGKN-----LKVKRDGKNIT-----FDLAKNL 793

Qy    1237 IIR-GLLNDTI--------------PQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSG 1281
              ::   ::||:              |::     :| |   ||           | ||
Db     794 EVKTAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFAKET-----------ADASG 841

Qy    1282 TQHEAVPG------DDNDGSTSSVSSYDV-------VTNVRDV-----GISTNGEVTGEE 1323
             :::  : |      : : |: ||    :|         :: ||      |  ||
Db     842 SKNVYLKGIATTLTEPSAGAKSSHVDLNVDATKKSNAASIEDVLRAGWNIQGNGNNVDYV 901

Qy    1324 ETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGE 1383
             |:   :| :|  ::    :     :: |  |:           || |      :  :|:
Db     902 ATYD--TVNFT--DDSTGTTTVTVTQKADGKGA----------DVKIGAKTSVIKDHNGK 947

Qy    1384 --TAREV----------EVDGK----GLLPESVVGEAPTQE----RGRAADGNTAQTAVN 1423
               | :::          | |||    ||:    | :|  :       |  |   |  ||||
Db     948 LFTGKDLKDANNGATVSEDDGKDTGTGLVTAKTVIDAVNKSGWRVTGEGATAETGATAVN 1007

Qy    1424 EGDREPVQS-SLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEK----PAP 1478
             |: | | | : |:     :    | |: :    || :|    |      | |:|
Db     1008 AGNAETVTSGTSVNFKNGNATTATVSKDNGNINVKYDV---NVGDGLKIGDDKKIVADTT 1064

Qy    1479 SVTTRGVKIIDKGKAVAHVAEKKQVQVE------------------QPKQRSLTINEGK 1519
             ::|  | |:     | :    || |  |                      :   |  | |
Db     1065 TLTVTGGKVSVPAGANSVNNNKKLVNAEGLATALNNLSWTAKADKYADGESEGETDQEVK 1124

Qy    1520 AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNG-- 1577
             || ::    |  |  | |       |     ::  | | | :::       |    |
Db     1125 AGDKV----TFKAGKNLKVKQSEKDFT-----YSLQDTLTGLTSITLGGTANGRNDTGTV 1175

Qy    1578 ----------------GSHVSSG-----------WPRALEDILTAIKYPSVFDHCLVQK 1609
                           |:  |:|                : : :: :|:|
Db     1176 INKDGLTITLANGAAAGTDASNGNTISVTKDGISAGNKEITNVKSALK----------T 1224

Qy    1610 YKMGGGVPFHADDEECYPSDNPILTVNLVGK--ANFSTKCRKGGK-VMVINVAS---GDY 1663
             ||          |:| : :       |  |||  |  | |    ||  : |:||    ||
Db     1225 YKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGD- 1283

Qy    1664 FLMPCGFQRTHLHSVN-SID--EGRISLTFRATRRVFGVGRMLQLAGG-----VSDEKSP 1715
             | ::     :   :|  :|   ||  ||:          :| |      :|  :
Db     1284 -----GLEKDTDGKIKLKVDNTDGNNLLTVDATKGA-------SVAKGEFNAVTTDATTA 1331

Qy    1716 GVPNQQPQ-------SQGATRTITPKSG-------GKALS------------------- 1741
             |   :        | ||| | | |           ||::
Db     1332 QGTNANERGKVVVKGSNGATATETDKKKVATVGDVAKAINDAATFVKVENDDSATIDDSP 1391
```

```
Qy   1742 --EGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAV 1799
Db   1392 TDDGANDALKAXDTLTLKAGKNLKVKRD---GKNITFALANDLS------VKSATVSDKL 1442

Qy   1800 VEYLKYLAIGIERTYRALLMARN-------------IAVTTAEGVLKVPNQVYESLPGFH 1846
Db   1443 SLGTNGNKVNITSDTKGLNFAKDSKTGDDANIHLNGIASTLTDTLLN---------SGAT 1493

Qy   1847 VYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAV-VALGDNLFVCDDILVFH 1905
Db   1494 TNLGGNGITDNEKKXAASVKD----------VLNAGWNVRGVKPASANNQVENIDFVATY 1543

Qy   1906 DAINLIGALKVARCGMV--GESFKSFEYKCYNAPPGGGKTTTLVD---------EFVKSP 1954
Db   1544 DTVDFVSGDKDTTSVTVESKDNGKRTEVKI------GAKTSVIKDHNGKLFTGKELKDAN 1597

Qy   1955 NSTATITANVGSSED------------INMA---VKKRDPNLEGLNSATTVNSRVVNF 1997
Db   1598 NNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTTGANGQNDDFATVASGTNVTF 1655


RESULT 13
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
;  GENERAL INFORMATION:
;    APPLICANT:  St. Geme III, Joseph W.
;    APPLICANT:  Barenkamp, Stephen J.
;    TITLE OF INVENTION:  HAEMOPHILUS ADHESION PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  Four Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/685,467
;      FILING DATE:  22-JUL-1996
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/409,995
;      FILING DATE:  24-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-61053-2/RFT/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1912 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-685-467-4


  Query Match             1.2%;  Score 138.5;  DB 3;  Length 1912;
  Best Local Similarity   18.1%;  Pred. No. 0.027;
  Matches  315;  Conservative  208;  Mismatches  598;  Indels  617;  Gaps   77;

Qy    627 GNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSFS-------------------NVS 666
Db    168 GNGNVHLNGLDST-------LPDAVTNTGVLSSSSFTPNDVEKTRAATVKDVLNAGWNIK 220

Qy    667 GFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFI 726
Db    221 G------AKTAGGNVESVDL------------------------------------------ 234

Qy    727 PVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITG---- 782
Db    235 -VSAYNNVEF--ITGDKNTLDVVLTAKENXKTT--EVKFTPKTSVIKEKDGKLFTGKENN 289

Qy    783 --SRVVHSKVDISPDDMWGLVVAVMAQAIKD-------RAKSIRSYN------------- 820
Db    290 DTNKVTSNTATDNTDEGNGLVT---AKAVIDAVNKAGWRVKTTTANGQNGDFATVASGTN 346

Qy    821 -FIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPG 879
Db    347 VTFESGDGTTASVTK-------DTNGNGITVKYDAKVGDGLKFDSD---------KKIVA 390

Qy    880 SVVVTICTSGASDRLELRGAFDISKETFGRKLKNS----------------------RLR 917
Db    391 DTTALTVTGGK--------VAEIAKEDDKKKLVNAGDLVTALGNLSWKAKAEADTDGALE 442

Qy    918 VFSRAIVEDSIKVMKAMKTEDGKPLPITED------------SVYAFIMGNVSN----- 959
Db    443 GISK---DQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSITLGGTTNGGNDA 499

Qy    960 --------VHCTRAGLLGGSKATVVSSVSKGLVARG--AATKAFSGITSFFSTGSLFYDR 1009
Db    500 KTVINKDGLTITPAG-NGGTTGTNTISVTKDGIKAGNKAITNVASGLRAY---------- 548

Qy   1010 GLTEDERLDALVRTENAINSPV-----GILETSRVAVSK--------------------V 1044
Db    549 ---DDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLRKLGWV 605

Qy   1045 VA---GTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGS 1101
Db    606 VSTKNGTKEESNQVKQADEVLF-----------TGAGAATVTSKSEN----------GK 643

Qy   1102 SYETLSSLSSQAAGGLRG-----------------LT-----SSTVSGGSLVVRRGFSSA 1139
Db    644 HTITVSVAETKADCGLEKDGDTIKLKVDNQNTDNVLTVGNNGTAVTKGGFETVKTGATDA 703

Qy   1140 ----VTVTRAT--------VAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFG 1187
Db    704 DRGKVTVKDATANDADKKVATVKDVATAINSAATFVKTE--------------------N 743

Qy   1188 LGTLL-----------GARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSL 1236
Db    744 LTTSIDEDNPTDNGKDDALKAGDTLTFKAGKN-----LKVKRDGKNIT-----FDLAKNL 793

Qy   1237 IIR-GLLNDTI--------------PQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSG 1281
Db    794 EVKTAKVSDTLTIGGNTPTGGTTATPKVNITSTADGLNFAKET-----------ADASG 841

Qy   1282 TQHEAVPG------DDNDGSTSSVSSYDV-------VTNVRDV-----GISTNGEVTGEE 1323
Db    842 SKNVYLKGIATTLTEPSAGAKSSHVDLNVDATKKSNAASIEDVLRAGWNIQGNGNNVDYV 901

Qy   1324 ETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGE 1383
Db    902 ATYD--TVNFT--DDSTGTTTVTVTQKADGKGA----------DVKIGAKTSVIKDHNGK 947

Qy   1384 --TAREV----------EVDGK----GLLPESVVGEAPTQE----RGRAADGNTAQTAVN 1423
Db    948 LFTGKDLKDANNGATVSEDDGKDTGTGLVTAKTVIDAVNKSGWRVTGEGATAETGATAVN 1007

Qy   1424 EGDREPVQS-SLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEK----PAP 1478
Db   1008 AGNAETVTSGTSVNFKNGNATTATVSKDNGNINVKYDV---NVGDGLKIGDDKKIVADTT 1064

Qy   1479 SVTTRGVKIIDKGKAVAHVAEKKQVQVE------------------QPKQRSLTINEGK 1519
Db   1065 TLTVTGGKVSVPAGANSVNNNKKLVNAEGLATALNNLSWTAKADKYADGESEGETDQEVK 1124
```

```
Qy   1520 AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNG-- 1577
Db   1125 AGDKV----TFKAGKNLKVKQSEKDFT-----YSLQDTLTGLTSITLGGTANGRNDTGTV 1175

Qy   1578 ---------------GSHVSSG-----------WPRALEDILTAIKYPSVFDHCLVQK 1609
Db   1176 INKDGLTITLANGAAAGTDASNGNTISVTKDGISAGNKEITNVKSALK----------T 1224

Qy   1610 YKMGGGVPFHADDEECYPSDNPILTVNLVGK--ANFSTKCRKGGK-VMVINVAS---GDY 1663
Db   1225 YKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNGKHTVTIDVAEAKVGD- 1283

Qy   1664 FLMPCGFQRTHLHSVN-SID--EGRISLTFRATRRVFGVGRMLQLAGG-----VSDEKSP 1715
Db   1284 -----GLEKDTDGKIKLKVDNTDGNNLLTVDATKGA-------SVAKGEFNAVTTDATTA 1331

Qy   1716 GVPNQQPQ-------SQGATRTITPKSG-------GKALS------ ------------- 1741
Db   1332 QGTNANERGKVVVKGSNGATATETDKKKVATVGDVAKAINDAATFVKVENDDSATIDDSP 1391

Qy   1742 --EGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAV 1799
Db   1392 TDDGANDALKAXDTLTLKAGKNLKVKRD---GKNITFALANDLS------VKSATVSDKL 1442

Qy   1800 VEYLKYLAIGIERTYRALLMARN-------------IAVTTAEGVLKVPNQVYESLPGFH 1846
Db   1443 SLGTNGNKVNITSDTKGLNFAKDSKTGDDANIHLNGIASTLTDTLLN---------SGAT 1493

Qy   1847 VYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAV-VALGDNLFVCDDILVFH 1905
Db   1494 TNLGGNGITDNEKKXAASVKD----------VLNAGWNVRGVKPASANNQVENIDFVATY 1543

Qy   1906 DAINLIGALKVARCGMV--GESFKSFEYKCYNAPPGGGKTTTLVD---------EFVKSP 1954
Db   1544 DTVDFVSGDKDTTSVTVESKDNGKRTEVKI------GAKTSVIKDHNGKLFTGKELKDAN 1597

Qy   1955 NSTATITANVGSSED------------INMA---VKKRDPNLEGLNSATTVNSRVVNF 1997
Db   1598 NNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTTGANGQNDDFATVASGTNVTF 1655
```

```
RESULT 14
US-08-232-537-2
; Sequence 2, Application US/08232537
; Patent No. 5516655
;  GENERAL INFORMATION:
;    APPLICANT:  Peery, Robert B.
;    APPLICANT:  Skatrud, Paul L.
;    TITLE OF INVENTION:  MULTIPLE DRUG RESISTANCE GENE OF
;    TITLE OF INVENTION:  AUREOBASIDIUM PULLULANS
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,537
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X9212
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;      TELEFAX:  317-276-1917
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1302 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-232-537-2
```

Query Match 1.2%; Score 135.5; DB 1; Length 1302;
Best Local Similarity 18.6%; Pred. No. 0.024;
Matches 203; Conservative 147; Mismatches 413; Indels 329; Gaps 45;

```
Qy    558 YVSSDLLTEAASKSVSYCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVF-DI 616
Db    360 FLGSFATAAAAGESVYEILNHPQS---------EINVYSEAGQEATES-DMKADLVFRNV 409

Qy    617 ALMFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQCGE---------MYEHSFSNVSG 667
Db    410 TFVYPAR----TSARALEEMSLILKAGQMNAIVGTSGCGKSTLVSLLLRLYDIS----SG 461

Qy    668 FFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPS----FVGKSL 723
Db    462 QLTIGSHDIKDFNVRSLR-----------------------KYTALVDQDSVLFSGSVL 497

Qy    724 VFIPVVAGSSVSFKTIVLDS---------DFVDRIYSYALNTIGTFENRTFEYAVGAVR 773
Db    498 ENISYGLGEHSLSDDVVLERCTEAAKAANLDFVDFLPQGIHTRIG---NGGYTSLSG--- 551

Qy    774 SQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVF 833
Db    552 GQNQRICLARALVKKPALLLLDEPTAALDANSEGLIMDAVKSV-------AATGTTVVMV 604

Qy    834 KLFFQTVGDCFSNAVSVYAKAMV----HDNFNVLETLMSMPRAFIRKVPGSVVVTICTSG 889
Db    605 AHRLSTVSDS-PNIVLMGAGKVIEQGNHDE------LMQLEGAYFNLIQAQQLNDADESS 657

Qy    890 ASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSV 949
Db    658 A----------EVSAATTSQVTPQKASKSEDSAASSDTETVPPQAKKED---KPAKKAGF 704

Qy    950 YAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDR 1009
Db    705 WKLLL------RCLR-----------LAKSDSPIIALGLAASIVSG-------------- 733

Qy   1010 GLTEDERL--DALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLR- 1066
Db    734 GIILGEAIVFGNLISVLNDLESP-----------------DFRSRADLFSLLFFILAL 774

Qy   1067 --------NKVLIGI----FVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAA 1114
Db    775 IALFSYAGNGCCFGIVSSHFVAKIQHISLASILRQ----DMQWFSGQSVPSLMSSLSSDA 830

Qy   1115 GGLRGLTSSTVSGGSLVVRRGFSSAVTVT----RATVAKRQVPLALLSFSTSYAISGCSM 1170
Db    831 GQLACL--SGVAIGTI-----FTVCVSITGGIILAHVVAWKIAVVLLAAVPVMITAGYVR 883

Qy   1171 LGIWAHALPRHLMFFFGLGTLL-----GARASANTWKFGGFSN-NWCAVPEVVWRG---K 1221
Db    884 LRVLALAESRHRSAYNDAASIAAEACRGIRTIASLGRERGVSRASNAAVKEPYDKGIRFT 943

Qy   1222 SVSSLLLPITLGVSLIIRGLL---------NDTIPQLAY--VPP---------------- 1254
Db    944 LITNTLLALSFSITYFVYALAYWWGAKQVRNGTYSQLDFFIVLPALLFSAQSAGQIFSLS 1003

Qy   1255 -------VEGRNVYDETLRYYRDFDYD---EGAGPSGTQHEAVPGDDNDGSTSSVSSYDV 1304
Db    1004 PEMSRAGVAARNVFGLHDQKPTIVDVDAKQSGALPSSTL--SIPTLEDKASPSSGGWIEF 1061

Qy   1305 VTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAP---SAAVAERQGD------PSG 1355
```

```
Db   1062 ----------------------KNVSLCYPSKPQHPALQNVNISIRPGEFIALVGPSG 1097

Qy   1356 SGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGK-------------GLLPESV- 1400
Db   1098 AGKSTILSLL------------QRFYDPTAGSVQLDGQDIREVAVPQHRGRLGLVPQEPD 1145

Qy   1401 -----------VGEAPTQ----------------------ERGRAADGNTAQTAVNEGD 1426
Db   1146 LFPGSISYNIGLGAAPGQLVTRDDIEKICAKCGIHEFIMSLPEGYSTECGTNGSKLSGGQ 1205

Qy   1427 RE--PVQSSLVSSPQADI--PKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTT 1482
Db   1206 KQRIAVARALIRSPEVLLLDEYTSALDAHSEQQIKEAVDGASVDRTTIVVAHRLSTVQNA 1265

Qy   1483 RGVKIIDKGKAV 1494
Db   1266 DRIFVFDDGRVV 1277


RESULT 15
PCT-US95-10661A-5
; Sequence 5, Application PC/TUS9510661A
;  GENERAL INFORMATION:
;    APPLICANT:  Washington University, et al.
;    TITLE OF INVENTION:  Haemophilus Adherence and Penetration Protein
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;     ·ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/10661A
;      FILING DATE:  16-AUG-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/296,791
;      FILING DATE:  25-AUG-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  FP-59941/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1702 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
PCT-US95-10661A-5


  Query Match             1.2%;  Score 135;  DB 4;  Length 1702;
  Best Local Similarity   19.9%;  Pred. No. 0.045;
  Matches  181;  Conservative   97;  Mismatches  340;  Indels  292;  Gaps   42;

Qy    721 KSLVFIPVVAGSSVSFKTIVLDSD-FVDRIYSY------ALNTIGTFENRTFEYAVGAVR 773
Db    492 KAFSQVGIVSGRS----TVVLNDDKQVDPNSIYFGFRGGRLDANG--NNLTFEHI----- 540

Qy    774 SQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVF 833
Db    541 ---RNIDDGARLVNHNTSKTS------TVTITGESLITDPNTITPYNIDAPDED------ 585

Qy    834 KLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDR 893
Db    586 ---------------NPYAFRRIKDGGQLYLNLENYTYYALRK-----------GASTR 618

Qy    894 LEL--------------RGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDG 939
Db    619 SELPKNSGESNENWLYMGKTSDAAKRNVMNHINNERMNGFNGYFGE----------EEG 667

Qy    940 KPLPITEDSVYAFIMGNVSNVHCTRAG--------LLGGSKATVVSSVSKG-LVARGAAT 990
Db    668 K---------------NNGNLNVTFKGKSEQNRFLLTGGTNLNGDLKVEKGTLFLSGRPT 712

Qy    991 KAFSGITSFFSTGSLFYDRGLTEDERL----DALVR---------TENA----------I 1027
Db    713 PHARDIAGISSTKK---DQHFAENNEVVVEDDWINRNFKATNINVTNNATLYSGRNVANI 769

Qy   1028 NSPVGILETSRVAV--------------SKVVAGTKEFWSEVSLNDF------------- 1060
Db    770 TSNITASDNAKVHIGYKAGDTVCVRSDYTGYVTCTTDKLSDKALNSFNATNVSGNVNLSG 829

Qy   1061 -TTFVLRNKVLIGIFVASLGAAPIA------WKY------------RRGIAANARRYAG- 1100
Db    830 NANFVLGKANLFGT-ISGTGNSQVRLTENSHWHLTGDSNVNQLNLDKGHIHLNAQNDANK 888

Qy   1101 -SSYETLSSLSSQAAGGLRGLTS-STVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLS 1158
Db    889 VTTYNTLTVNSLSGNGSFYYLTDLSNKQGDKVVVTKSATGNFTLQVADKTGEPTKNELTL 948

Qy   1159 FSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKF------GGFSNNWCA 1212
Db    949 FDASNATR--NNLNV----------------SLVGNTVDLGAWKYKLRNVNGRYD---LY 987

Qy   1213 VPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQ--LAYVPPVEGRNVYDETLRYYRD 1270
Db    988 NPEVEKRNQTVDT----------------TNITTPNNIQADVPSVPSNN--EEIARVETP 1029

Qy   1271 FDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRS 1330
Db   1030 VPPPAPATPSETTETVAENSKQESKTVEKNEQDATETT-----AQNGEVAEEAKP----S 1080

Qy   1331 VQYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEV 1390
Db   1081 VKANTQTNEVAQSGSETEETQTTEIKETA-------KVEK-----------EEKAKVEK 1121

Qy   1391 DGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEV 1450
Db   1122 EEKAKVEKDEIQEAPQM---------ASETSPKQAKPAPKEVSTDT-------KVEETQV 1165

Qy   1451 HAQKEVKQEVPLATVSGATPIVDEKPA----PSVTTRGVKIIDKGKAVAHVAEKKQVQ-- 1504
Db   1166 QAQPQT-QSTTVAAAEATSP--NSKPAEETQPSEKT-------NAEPVTPVVSKNQTENT 1215

Qy   1505 VEQPKQRSLT 1514
Db   1216 TDQPTEREKT 1225


Search completed: January 31, 2001, 15:27:41
Job time: 110 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:27:32 ; Search time 60.61 Seconds
(without alignments)
2506.084 Million cell updates/sec

Title: US-09-301-906-15
Perfect score: 11520
Sequence: 1 MDYIRPLRVFSFPHVNNTLE.....QSVSDALLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_66:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	629	5.5	1729	2	A49282	fusion protein 1a/
2	402.5	3.5	1547	2	JQ0096	hypothetical 176K
3	400	3.5	1718	1	JQ1734	genome polyprotein
4	391	3.4	1707	2	S77908	hypothetical prote
5	389	3.4	1707	2	S77910	hypothetical prote
6	338.5	2.9	1707	2	S01865	genome polyprotein
7	329.5	2.9	1126	1	WMFM12	125K protein - alf
8	309	2.7	1615	1	WMTM8T	180K protein - tom
9	304	2.6	1766	2	S03701	141K protein - pea
10	297	2.6	1611	1	WMTMPV	183K protein - pep
11	294.5	2.6	1646	1	WMTMS2	186K protein - cuc
12	291	2.5	1694	2	JQ1896	hypothetical 191.1
13	286.5	2.5	966	1	P1BVBB	RNA 1a protein - b
14	277	2.4	993	2	JA0074	RNA 1 protein - cu
15	275.5	2.4	1005	1	P1VXPJ	RNA 1 protein - pe
16	273	2.4	993	1	P1VXCM	RNA 1 protein - cu
17	272.5	2.4	1608	1	WMTMGM	183K protein - tob
18	271.5	2.4	1139	1	PAVBBS	alpha-a protein -
19	270.5	2.3	961	1	P1BVA	RNA 1a protein - b
20	269	2.3	993	2	E71392	ATP-dependent RNA
21	267	2.3	1615	1	WMTM18	183.3K protein - t
22	264.5	2.3	1616	1	JQ2144	183K protein - tom
23	264	2.3	1116	2	S26358	hypothetical prote
24	263	2.3	991	1	P1BVCV	RNA 1 protein - cu
25	259.5	2.3	1885	2	JQ2183	hypothetical 216.5
26	253.5	2.2	993	2	JQ2169	RNA 1 protein - cu
27	252	2.2	958	1	P1BVCC	1a protein - cowpe
28	252	2.2	1112	2	S49432	replicase 126K - o
29	245.5	2.1	1884	1	A45353	genome polyprotein
30	237	2.1	993	1	P1VXTA	RNA 1 protein - to
31	236	2.0	1601	2	S48699	178K protein - tob
32	218	1.9	1597	2	S65053	genome polyprotein
33	198	1.7	1589	2	T42233	submaxillary mucin
34	193	1.7	2059	2	D82671	surface protein XF
35	187	1.6	7962	2	I38346	elastic titin - hu
36	179	1.6	345	2	S23862	hypothetical prote
37	173.5	1.5	1848	2	A44140	cellulose-binding
38	171	1.5	1691	1	A44212	genome polyprotein
39	169.5	1.5	1693	1	MNWWHE	genome polyprotein
40	165.5	1.4	4152	2	T31102	filamentous hemagg
41	165.5	1.4	4919	2	T31105	hypothetical prote
42	164.5	1.4	1088	3	T41591	hypothetical serin
43	164.5	1.4	1131	2	T41144	hypothetical serin
44	164.5	1.4	1608	2	A28182	hemolysin A - Serr
45	163.5	1.4	3029	2	S76109	hypothetical prote

ALIGNMENTS

RESULT 1
A49282
fusion protein 1a/1b - sugar beet yellows virus (isolate BYV-U) (fragments)
C;Species: sugar beet yellows virus, SBYV
C;Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 17-Mar-1999
C;Accession: A49282
R;Agranovsky, A.A.; Koonin, E.V.; Boyko, V.P.; Maiss, E.; Froetschl, R.; Lunina, N.A.
Virology 198, 311-324, 1994
A;Title: Beet yellows closterovirus: complete genome structure and identification of
A;Reference number: A49282; MUID:94082464
A;Accession: A49282
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-1729 <AGR>
A;Cross-references: GB:X73476
A;Note: this protein is predicted to result from +1 ribosomal frameshifting during tr

Query Match 5.5%; Score 629; DB 2; Length 1729;
Best Local Similarity 20.8%; Pred. No. 6e-31;
Matches 388; Conservative 201; Mismatches 525; Indels 748; Gaps 76;

```
Qy    429 LLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPREREKLRELFPELSIQFSDSVRSSHP 488
          |   ::  | |   :  :: :|| |     ::    :  |   :|: :: |: || | ||
Db      4 LARSVEKDLIDFKDEIKSLSKEKRSVTVPFYMGEAVQSGLTRAYPQFNLSFTHSVYSDHP 63

Qy    489 FANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKRRINEIL 548
          |   |   |   :    :  | ||||   :|:| |  : ||| |: |:|| :||::
Db     64 AAAGSRLLENETLASMAKS-SFSDIGGCPLFHIKRGSTDYHVCRPIYDMKDAQRRVSR-- 120

Qy    549 FLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDK 608
          |   |    :| : | || :: || |     ||: ::|   || ||| |  ::| ||
Db    121 ELQARGLVENLSREQLVEAQAR-VSVCPHTLGNCNVKSDVLIMVQVYDASLNEIASAMVL 179

Qy    609 KGALVFDIALMFPVELLYGNGEVYLEEL--DTLVKREGDYLAYNVGQCGEMYEHSFSNVS 666
          | : |  : :: | |||       :: |  | :|    | : |  |     | |  ||:
Db    180 KESKVAYLTMVTPGELLDEREAFAIDALGCDVVVDTRRDMVQYKFG--SSCYCHKLSNIK 237

Qy    667 GF-----FTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPG------------TE 708
              |||       |||:| :|    | | ::  : |:  ||                ||
Db    238 SIMLTPAFTF------SGNLFSVEMYENRMGVNYYKITRSAYSPEIRGVKTLRYRRACTE 291

Qy    709 VTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYA 768
          |    | | |      |:          :  | ||: || |:: | ::      ::|||:
Db    292 VVQVKL-PRFDKTLKTFLS-------GYDYIYLDAKFVSRVFDYVVSNCSVVNSKTFEWV 343

Qy    769 VGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDR------AKSIRSYNFI 822
             ::| |: |:   :|:|  | |           ::|  :: |      ||::  |
Db    344 WSYIKSSKSRVVISGKVIHRDVHIDLKHSECFAAVMLAVGVRSRTTTEFLAKNLNYY--- 400

Qy    823 KASEGSLAGVFKLFFQTVGD--CFS--NAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVP 878
                          ||  ||:   |:| |:
```

```
Db    401 -----------------TGDASCFALRRALSAYS------------------------- 417

Qy    879 GSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTED 938
                                          || ||   | |:|
Db    418 ---------------------------------NSALR---RKIIE------------- 427

Qy    939 GKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITS 998
                     || ||: :
Db    428 -------------FIFGNIHHP----------------------------------- 436

Qy    999 FFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLN 1058
             :|                        | ::||: ||        :    || ::
Db    437 --------FD---------------------VAVIETNEVA-------PEPLSPEVDID 459

Qy   1059 ---DFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSS--LSSQA 1113
            ||                                          ||  |::||  ::|
Db    460 VDCDF-------------------------------------GSDSESVSSDEVASNP 480

Qy   1114 AGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAK--RQVPLALLSFSTSYAISGCSML 1171
            ||         |||    |  |: :|     | |  |::|  |       |
Db    481 RPGLH--------GGS----RRSSNFLTSLVKVVFKLARRIPRLLFRLRNFVAYF----- 523

Qy   1172 GIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPIT 1231
             :      : |  | ||  |           |  ||     :  ||:  :   |:|
Db    524 -VERRLASKRLKTFIGLARL----------FDNFS-----LTSVVYLLQEYDSVLNAF- 565

Qy   1232 LGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDD 1291
           : | ||   |||
Db    566 IDVELI---LLN------------------------------------------------ 574

Qy   1292 NDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQG 1351
             : ||:   :|: ||       | :|   |                     |:
Db    575 ----SGSVNVLPLVSWVR-------GSLTKLAE-------------------AIV---- 599

Qy   1352 DPSGSGTADAMAFVESVKKGVDDVFHQQSSG--ETAREVEVDGKGLLPESVVGEAPTQER 1409
             ||| |   :|:  :   | |     |:        |   || : | |    |   ||
Db    600 ---GSGFA---SFLGRMCCRVSDWCSSSSNAGCNFMSPVRTKGKFVPPSSSGSTASMYER 653

Qy   1410 GRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGAT 1469
           | : :          || | |:        :|   |
Db    654 LEALESDI---------REHVLSTC---------RVGSDE-------------------- 675

Qy   1470 PIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRT 1529
                                          : |:||:
Db    676 ---------------------------------EEERPKE------------------ 682

Qy   1530 CSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRAL 1589
                                                  |  |  | |
Db    683 ---------------------------------------VTEPGIEHTS------- 692

Qy   1590 EDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRK 1649
           ||::      :  |      : ||   :::| |    :|         :||
Db.   693 EDVV------PIRSH----SQPLSGGECSYSEDRE----ENE--------RAN------- 723

Qy   1650 GGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGV 1709
                        |:|      |:  : |   | :    |  | : ||   |
Db    724 --------------LLP------HVSKIVSERRG-LETARRNKRTLHGVSEFLNAINTS 761

Qy   1710 SDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGRE---VKGRSTYSIWCE-QDYVRK 1765
           :::  | : :  |:|:  | ::              ||   ::  : : : |: ::|  :
Db    762 NEQPRPIIVDHSPESRALTNSV-------------REFYYLQELALFELSCKLREYYDQ 807

Qy   1766 CEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAV 1825
            :    :      : |              ||  |                        :
Db    808 LKVANFNRQECLCDKD-------------EDMFV----------------------L 829

Qy   1826 TTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDV 1885
             :||:   |     ||  |    | :  | |   ||    :||
Db    830 RAGQGVVSGRN---SRLPLKHF--KGHEFCFRS--GGL----VPY--------------- 863

Qy   1886 DAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYK-------CYNAPP 1938
                |     | |  ||   | : |      | :  |:::| :         | |||
Db    864 -------DGTSRVDTI--FHTQTNFVSA-NALLSGYL--SYRTFTFTNLSANVLLYEAPP 911
```

```
Qy   1939 GGGKTTTLVDEFVK--SPNSTATITANVGSSEDI-----NMAVKKRDPNLEGLNSATTVN 1991
            |||||||:  | :  |  ::  :|||  | |:|      : : : |  |:  :   |::
Db    912 -GGKTTTLIKVFCETFSKVNSLILTANKSSREEILAKVNRIVLDEGDTPLQTRDRILTID 970

Qy   1992 SRVVNFIVRGMYKRVL-VDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFR 2050
           | ::|   ||:  :|| :||  |:| |      :  |      : |||  || :    :
Db    971 SYLMN--NRGLTCKVLYLDECFMVHAGAAVACIEFTKCDSAILFGDSRQIRY---GRCSE 1025

Qy   2051 MDCAV------FVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVV------S 2098
           :| ||      ||   || ||   |||||  |||  ||:         ||: |       :
Db   1026 LDTAVLSDLNRFV-DDESRVYGEVSYRCPWDVCAWLSTF--------YPKTVATTNLVSA 1076

Qy   2099 GKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDM-----KRSLKGKGKETPVMT 2153
           |:      |:  | | : |||   :  ||| | | || |:     |||     | | |:|
Db   1077 GQS-----SMQVREIESVDDVEYSSEFVYLTMLQSEKKDLLKSFGKRSRSSVEKPT-VLT 1130

Qy   2154 VHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYIS 2213
           ||||||:|:  | | |||  :|  |  : || | ||||  || |: |||| || :   |
Db   1131 VHEAQGETYRKVNLVRTKFQEDDPFRSENHITVALSRHVESLTYSVLSSKRDDAIAQAIV 1190

Qy   2214 DA 2215
            |
Db   1191 KA 1192
```

```
RESULT   2
JQ0096
hypothetical 176K protein - papaya mosaic virus
C;Species: papaya mosaic virus
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Jun-2000
C;Accession: JQ0096
R;Sit, T.L.; Abouhaidar, M.G.; Holy, S.
J. Gen. Virol. 70, 2325-2331, 1989
A;Title: Nucleotide sequence of papaya mosaic virus RNA.
A;Reference number: JQ0096; MUID:89381685
A;Accession: JQ0096
A;Molecule type: genomic RNA
A;Residues: 1-1547 <SIT>
A;Cross-references: GB:D13957; GB:D00580; NID:g222330; PIDN:BAA03050.1; PID:g222331
A;Note: this protein shares two major domains of similarity with other potexvirus pro
C;Comment: The genome is a single-stranded, positive-sense RNA.
C;Superfamily: eggplant mosaic virus RNA-directed RNA polymerase


  Query Match             3.5%;  Score 402.5;  DB 2;  Length 1547;
  Best Local Similarity   24.0%;  Pred. No. 1.2e-16;
  Matches  177;  Conservative   87;  Mismatches  230;  Indels  243;  Gaps   33;

Qy   1432 SSLVSSPQADIPKVTQSEVHAQK-EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDK 1490
           | |   |:    |  ::    |: |||:|   |: ||    : | ||              |
Db    466 SLLQKEPEVMEAKEAEATDEPQRPEVKEEQAEASTSGRAEEIQEDPATK----------K 515

Qy   1491 GKAVAHVAEKKQVQVEQPK-QRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFS 1549
           ||             |:|   | |                 | ||: | : |       |
Db    516 GK-------------EEPNPNRDL--------------LCPCGLHLKIKN-----AEFP 542

Qy   1550 N--AFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTA--IKYP-SVFDH 1604
                    |:| || | |||| |: |:| |||| | |||  || || |  || |   |:
Db    543 ELPVLDHPDHLTGRKAWFFSKDGKPYSYTGGSHASRGWPNWLEKILAAIEIKEPLPEFNQ 602

Qy   1605 CLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYF 1664
           ||||::|:   :||| ||| |||  : :||:|  |:       |:||     | :  |||:
Db    603 CLVQQFKLQAAIPFHRDDEPCYPKGHQVLTINHSGECLTQIACQKG--KASITMGFGDYY 660

Qy   1665 LMPCGFQRTHLHSVNSIDEGRISLTFRAT--------------------------RRVF 1697
           | | ||| :| |:|::   ||:|||||                           : :
Db    661 LSPVGFQESHKHAVSNTTGGRVSLTFRCTVQQNKFNDDGSMEALDNLPWKAWIPKLQNLG 720

Qy   1698 GVGRMLQ------LAGGVSDEKSPGVPNQQPQS---------QGATRTITPKSGGKALSE 1742
             || ||      |   : : :|  :|  :|:           |  |  ||  |      :
Db    721 FQGRQLQYDPNGALISPIEEIRS--MPKCKPEGVPEVVYKTLDGLARAPTPYSPNPIRAR 778
```

```
Qy   1743 GSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEY 1802
Db    779 AYTSDVKNCRIGALLRQQGK----EW-----------------------GCRFDALVEA 810

Qy   1803 LKYLAIGIERTYRALLMARNIAVTTAEGV-LKVPNQVYESLPGFHVYKSGTDLIFHSTQD 1861
Db    811 GK---------------RELAISVIHGAGGSGKSQALQTL------------------- 835

Qy   1862 GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGM 1921
Db    836 ---IKDNP-------------ELDITVVLPTN-----------ELRLDWLRKLPKAPQ 866

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPN----------STATITANVGSSEDIN 1971
Db    867 --EKFKTFE-KALLAPP---TPIVIFDDYGKLPAGYVEAFCLYFSTVQLIILTGDC---- 916

Qy   1972 MAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYK--RVLVDEVHMMHQGLL-QLGVFA--T 2026
Db    917 ----KQSVHHESNENATTSS---IEPLVKEASELCRYYINATHRNKKDLANKLGVYSEKT 969

Qy   2027 GASE---------GLFFGDINQIPFINREKVFR---------MDCAVFVPKKESVVYTSK 2068
Db    970 GLTEVTHGTTPIPGLHM----LVPSLYKKQAFSEMGHKVSTYAGCQGITAPKIQILLTEE 1025

Qy   2069 SYRCPLDVCYLLSSMTV 2085
Db   1026 TSLCSREVLYTALSRAV 1042
```

RESULT 3
JQ1734
genome polyprotein - shallot virus X
N;Alternate names: RNA nucleotidyltransferase (RNA-directed); RNA replicase
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: shallot virus X
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 11-Jun-1999
C;Accession: JQ1734
R;Kanyuka, K.V.; Vishnichenko, V.K.; Levay, K.E.; Kondrikov, D.Y.; Ryabov, E.V.; Zavriev
J. Gen. Virol. 73, 2553-2560, 1992
A;Title: Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal cistron closel
A;Reference number: JQ1734; MUID:93019008
A;Accession: JQ1734
A;Molecule type: genomic RNA
A;Residues: 1-1718 <KAN>
A;Cross-references: GB:M97264; NID:g295078; PIDN:AAA47787.1; PID:g295079
C;Superfamily: eggplant mosaic virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase; P-loop; RNA biosynthesis; RNA replication
F;915-922/Region: nucleotide-binding motif A (P-loop)
F;978-983/Region: nucleotide-binding motif B
F;921/Binding site: ATP (Lys) #status predicted

Query Match 3.5%; Score 400; DB 1; Length 1718;
Best Local Similarity 23.5%; Pred. No. 2e-16;
Matches 187; Conservative 114; Mismatches 290; Indels 206; Gaps 37;

```
Qy   1530 CSCGVQLDVYNEATIATRFSNAF-----TFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSG 1584
Db    583 CECGTEITV-------NSFGRAIEVAGVNLTDHMKGRLAAFYSRDGQGYSYTGYSHKSQG 635

Qy   1585 WPRALEDILTAI-KYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANF 1643
Db    636 WLEGLDKLIEACGEKPTTYNQCLVQKYEQGSRIGFHSDEQAIYPKGNKILTVNAAGSGTF 695

Qy   1644 STKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRML 1703
Db    696 GIKCAKG--ETTLNLEDGDYFQMPSGFQETHKHNVVAVTP-RLSFTFRST---------- 742

Qy   1704 QLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGRE------VKGRSTYSIW 1757
Db    743 ----VVNSQKKPAEPEKLNQNNACPKPSDP-------SNASGKQHKKTHPAKGNE----- 786

Qy   1758 CEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVK----------TGTSEDAVVEYLKYL 1806
Db    787 -----------KSSSP--NLEPLDAP-TVEILKLHGFTALTPQHDGTCQIRPVYFNK-- 829

Qy   1807 AIGIERTYRALLM---AR---NIAVTTAEGVL--KVPNQ--------VYESLPGFHVYKS 1850
Db    830 DIHLRRKAVKTDMSPPARPFFDLATSLHRGIYTHKIDNRRATAYMSDVKNNLTGLVLPKL 889

Qy   1851 GTDLIFHSTQDGL---RVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDA 1907
Db    890 DRDLL--SSWVALAETTTREVAVLAIHGAGGAGKSRALQELLRSSPEL--ADSINIVVPT 945

Qy   1908 INLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSS 1967
Db    946 INLANDWKAKLPQMDPRRVMTFQKACERE----CKSVTIFDDYGKLP--AGFVDAYLAIK 999

Qy   1968 EDINMAV-----------KKRDPNLEGLNSATTVNSRVVNFIVRGMY------------ 2003
Db   1000 VNVELAILTGDQRQSTHHQERESQISSLQSNIAQFSKYADYYLNATHRQPRRLANPIKVH 1059

Qy   2004 ----------KRVLVDEVHMM------HQGLL-QLGVFAT--GASEGLFFGDINQI---- 2040
Db   1060 AERQLGGAVLKANIVPDLAMVLVPAFRSQSLLTDLGRHAMTYAGCQGLTLNHLTIILDKD 1119

Qy   2041 -PFINREKVFRMDCAVFVPKKESVVYT-----SKSYRCPLDVCYLLSSMT--VRGTEKCY 2092
Db   1120 TPLCSDEVLY----TAFSRASESITFVNTHSDNPAFLAKLDATPYLKTLISWVREDEE-- 1173

Qy   2093 PEKVVSGKDKPVVRSLSKR-PIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPV 2151
Db   1174 -----AGADCPATEPLVKDVPTKTHIPVANDKVQLEGKIEAMEDKDTRELWSGEEKTNLM 1228

Qy   2152 MTVHEAQGKTFSDVVLFRTKKA-DDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGT 2210
Db   1229 QTQDPV-------VQLFPHQQAKDEALF-----------------------KI--TIGE 1255

Qy   2211 YISDASPQSVSDALLHT 2227
Db   1256 RIRMATPEQNAKQLRHT 1272
```

RESULT 4
S77908
hypothetical protein 1 - grapevine trichovirus B
C;Species: grapevine trichovirus B
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
C;Accession: S77908
R;Saldarelli, P.
submitted to the EMBL Data Library, March 1996
A;Reference number: S77908
A;Accession: S77908
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1707 <SAL>
A;Cross-references: EMBL:X75448; NID:g1220265; PIDN:CAA53196.1; PID:g1061119
C;Superfamily: grapevine trichovirus B hypothetical protein

Query Match 3.4%; Score 391; DB 2; Length 1707;
Best Local Similarity 20.1%; Pred. No. 7.4e-16;
Matches 218; Conservative 137; Mismatches 306; Indels 422; Gaps 41;

```
Qy   1315 TNGEVTGEEE---THSPRSVQYTYVEEEVAPS-------AAVAERQGDPSGSGTADAMAF 1364
Db    338 TNGEESTAEALFVAQLSRQIQETKLYDKMGNFSLKECIWSAVSSAMGDSLVYFFDKARFY 397

Qy   1365 VESVKKGV---------------DDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQER 1409
Db    398 NESLESFILNCGPAKIEIKRVYRDAVYRNQ--GVCPHLVSFDGS---PDGVMDELYLQH- 451

Qy   1410 GRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGAT 1469
Db    452 -----------IRDMDRSSEPYSLTSA------EVLESSIGCR---------------- 477
```

```
Qy   1470 PIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQ----------PKQRSLTINEG 1518
Db    478 --------------ISLIEMGRAMLKFHQQQDVMINERVVLSPLDDFRSMRKALTDRDG 522

Qy   1519 KAG-------------KQLCM-----------------------------FRTCSCGV 1534
Db    523 ILGTTCLLLEWEYDPLKRYSMGLITATELNAAIIGTTKITAKKEMAKECSERYQQCSCGV 582

Qy   1535 QLDVYNEAT---IATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALED 1591
Db    583 RLPMRKLETPLKVPTE-----EFTDALKGRRAAFYSRHSWNYSYTGANHASKGWPEWVSN 637

Qy   1592 ILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGG 1651
Db    638 IAEKLGLGEDFDHCLAQIYEEDKGINFNADDEPCY-TDPEVVTVNLNGNANFHLKC---G 693

Qy   1652 KVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFR-----------------ATR 1694
Db    694 SESVVPLSDGDVLIMPKGFQKTHKHAVTGTSAGRISLTFRNGINAPDEGSDKMSEYEETH 753

Qy   1695 RVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITP---------------------- 1733
Db    754 RIGG---------------KPGCPREEQKESVSLQILADHMRVDLAICTSMVFAKDPRAR 798

Qy   1734 ---KSGG----------KALSEGSGREVKGRSTY--------SIWCEQDYVRKCEWLRAD 1772
Db    799 EEVKRGGMTLGTFICVLKSLNLGAYIESERGCLYVNGAYKELSCYAEDDHISEWSGPRKE 858

Qy   1773 ---NPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAE 1829
Db    859 TTFSNALSMNPDIRVVKYEVCEQ-------------------------RAKKLVDSFQE 892

Qy   1830 GVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKD-VDAV 1888
Db    893 GFTGV------CLNKFQKQKSSFNLV-------------------------NGKELIDVH 921

Qy   1889 VALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVD 1948
Db    922 LTLG----------------------------FAGSGKSFYPQC-------------VL 939

Qy   1949 EFVKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLV 2008
Db    940 KNSHYANSLVIVPRKALCSDWSN----KVHPDVKVVTFESAFRQQ------KKGYGLIVI 989

Qy   2009 DEVHMMHQGLLQL--GVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVV-- 2064
Db    990 DEIGLLPPGYIDMVHGYFCYDSL--LVLGDPLQCEYHSKADHF------FLGQEESVFKK 1041

Qy   2065 ------YTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDD 2118
Db   1042 FKGHCNYLYKSHRLPRN--QKLFEIECDGAE---GEGVTFNKPRA--------------- 1081

Qy   2119 VAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSL- 2177
Db   1082 -----KDLTLCASQ----------RRKDSEKDVNTVGESQGLSANRVNILLDK--DWSLV 1124

Qy   2178 --------FT---KQPHI--------------------LVGLSRHTRSLVYAALSSKLD 2205
Db   1125 NDETVIVAFTRARKEINIIGDASLVNNLKRSAKSTVLKKILGGERVTEGLILSLIRKKLP 1184

Qy   2206 DKV 2208
Db   1185 DVI 1187
```

RESULT 5
S77910
hypothetical protein 1 - grapevine virus A
C;Species: grapevine virus A
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 20-Jun-2000
C;Accession: S77910
R;Minafra, A.
submitted to the EMBL Data Library, June 1996

A;Reference number: S77910
A;Accession: S77910
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1707 <MIN>
A;Cross-references: EMBL:X75433; NID:g1405614; PIDN:CAA53182.1; PID:g1405615
C;Superfamily: grapevine trichovirus B hypothetical protein

Query Match 3.4%; Score 389; DB 2; Length 1707;
Best Local Similarity 21.3%; Pred. No. 9.9e-16;
Matches 181; Conservative 114; Mismatches 259; Indels 294; Gaps 35;

```
Qy   1426 DREPVQSSL----VSSPQADI-------PKVTQSEVHAQKEVKQEVPLATVSGATPIVDE 1474
Db    535 DIEDIQAGLEEGVITEEEAELRLLPTTKPKIT--EIHMD------------------DD 573

Qy   1475 KPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGV 1534
Db    574 TPG----TSGESDVEKFKSV--------------------------RSLC--------- 593

Qy   1535 QLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILT 1594
Db    594 REEIYSE---------------KLKGREVAFYSRHSKEYKYNGGSHRSLGWDEALNELTQ 638

Qy   1595 AIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVM 1654
Db    639 ELGLDDSYDHCLIQRYTAGGSIGFHADDEPCYLPGGSVVTVNLHGDATFEVKENQSGKIE 698

Qy   1655 VINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKS 1714
Db    699 KKELHDGDVYVMGPGMQQTHKHRVTSHTDGRCSITLR--------NKTVDYEARKGDEDS 750

Qy   1715 PGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRK--------- 1765
Db    751 EYEEDKAELDEGI--------------------------------DYLQKNQGNMCSLK 777

Qy   1766 --CEWLRADNP-VMALEPDYTPMTF-EVVKTGTSEDAVVEYLKYLAIGIERTYRALLMAR 1821
Db    778 AFADHMQLSTPSVIAIVNGASPQTLREIEDGGYSLATLVNLSKALDF------------ 824

Qy   1822 NIAVTTAEGVLKVP---NQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGI 1878
Db    825 PIAIHGERGYAETPGSYRRLHLKITSGHVE--------------------PFEGVTSKGG 864

Qy   1879 FTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKV--ARCGMVGESFKSFEYKCYNA 1936
Db    865 FRE------AMLLGDGVGV-------------GHFRVDKAKADRLAQSF-------YN- 896

Qy   1937 PPGGGKTTTLVDEF---------VKSPNSTATITANVGS-------------SEDINMAV 1974
Db    897 ----GNTGVLLGKYNKGKMHTGEIEEPKEVLTAFGFAGSGKSHWCQTILKHCSVEKVLVI 952

Qy   1975 KKRDPNLEGLNSATTVNSRVVNFIVRGM----YKRVLVDEVHMMHQGLLQLGVFATGASE 2030
Db    953 SPRKVLRDDWVAKISKKHRVVTFEVAFMDDYGCKDIVIDEIGLLPPGYIDLVIAAHQPRT 1012

Qy   2031 GLFFGDINQIPFINREKVFRMDCA---VFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRG 2087
Db   1013 LVLLGDPLQSTYHSKRDNVVLEASQEDVFNRVRGKLPYLCYSHRLPRN-CKLF-EIECMG 1070

Qy   2088 TEKCYPEKVVSG----KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
Db   1071 AES--EKRVVYRSNRLKDEPTI----------------------CATRAMKEE-----K 1100

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFR----TKKADDSLFTKQPHILVGLSRHTRSL---V 2196
Db   1101 GSG----WYTVSETQGLSFKSCLIYLDEHWAKKEDED-------VMVALTRSRGEIGIHV 1149

Qy   2197 YAALSSKL 2204
Db   1150 TPALKKKL 1157
```

```
RESULT   6
S01865
genome polyprotein - tobacco rattle virus (strain SYM)
N;Alternate names: RNA replicase
N;Contains: 134K protein; 194K protein; RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: tobacco rattle virus, TRV
C;Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 17-Mar-2000
C;Accession: S01865
R;Hamilton, W.D.O.; Boccara, M.; Robinson, D.J.; Baulcombe, D.C.
J. Gen. Virol. 68, 2563-2575, 1987
A;Title: The complete nucleotide sequence of tobacco rattle virus RNA-1.
A;Reference number: S01865; MUID:88034943
A;Accession: S01865
A;Molecule type: genomic RNA
A;Residues: 1-1707 <HAM>
A;Cross-references: EMBL:X06172
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: nucleotidyltransferase
F;1-1707/Product: 194K protein #status predicted <MAT2>
F;1-1187/Product: 134K protein #status predicted <MAT1>
F;1188/Region: opal stop codon

  Query Match             2.9%;  Score 338.5;  DB 2;  Length 1707;
  Best Local Similarity   18.0%;  Pred. No. 1.6e-12;
  Matches  334;  Conservative  209;  Mismatches  514;  Indels  801;  Gaps   72;

Qy    445 DAMIREKPS------HRCDVFLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFN 498
Db     50 DDMIRERKTKDPVIVHE---VLSQKEQNKLMEIYPEFNIVFKDDKNMVHGFAAAERKLQA 106

Qy    499 GIFSRRCGNVC-FFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDS 557
Db    107 LLLLDRVPALQEVDDIGGQWSFWVTRGEKRIHSCCPNLDIRDDQREISRQIFLTAIGDQA 166

Qy    558 YVSSDLLTE------------AASKSVSYCSRESQNCDSRADAGF-------------MV 592
Db    167 RSGKRQMSENELWMYDQFRKNIAAPNAVRCNNTYQGCTCR---GFSDGKKKGAQYAIALH 223

Qy    593 DVYDISPQQVAEAMDKKGALVFDIALMF-PVELLYGNGEVYLEELDTLVKREGDYLAYNV 651
Db    224 SLYDFKLKDLMATMVEKKTKVVHAAMLFAPESMLVDEG--------PLPSVDGYYMKKN- 274

Qy    652 GQCGEMYEHSFSNVSGFF---TFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTE 708
Db    275 ---GKIY-------FGFEKDPSFSYIHDWE------EYKKYLL--------------GKP 304

Qy    709 VTYRSLVPSFVGKSLVFIP-VVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEY 767
Db    305 VSYQ-------GNVFYFEPWQVRGDTMLFS-----------IYRIA-------------- 332

Qy    768 AVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEG 827
Db    333 --GVPR-----------------------------------------RSLSSQEYYR---- 346

Qy    828 SLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICT 887
Db    347 ---------------------------------------------RIYISRWENMVVVPI-- 361

Qy    888 SGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITED 947
Db    362 ------------FDLVEST--------------RELVKKDLFVEKQFM----------D 384

Qy    948 SVYAFIMG------NVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFS 1001
Db    385 KCLDYIARLSDQQLTISNV------------KSYLSSNNWVLFINGAAVK---------- 422

Qy   1002 TGSLFYDRGLTEDERLDALVRT---ENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLN 1058
Db    423 ------NKQSVDSRDLQLLAQTLLVKEQVARPV-----MRELREAILTETKPI---TSLT 468

Qy   1059 DFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLR 1118

Db    469 D----------VLGLISRKL------WK----------QFA----------NKIAVGGFV 492

Qy   1119 GLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHAL 1178
Db    493 GMV------GTLI---GFYPKKVLT----------------------------WAKDT 513

Qy   1179 PRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITL-----G 1233
Db    514 PNG--------------------------------PELCYENSHKTKVIVFLSVVYAIGG 541

Qy   1234 VSLIIRGLLNDTIPQLAYVPPVE-GRNVYD--ETLRYYRDFDYDEGAGPSGTQHEAVPGD 1290
Db    542 ITLMRRDIRDGLVKKLCDMFDIKRGAHVLDVENPCRYY--------------------E 580

Qy   1291 DNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQ 1350
Db    581 INDFFSSLYSA------------SESGET--------------------VLPDLSEVKAK 608

Qy   1351 GDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERG 1410
Db    609 SD----------KLLQQKKEIADEFLSAKFSNYSGSSVRTS-----PPSVVGSSRSGLGL 653

Qy   1411 RAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEV-KQEVPLATVSGAT 1469
Db    654 LLEDSNV--------------------------LTQARVGVSRKVDDEEIMEQFLSG-- 684

Qy   1470 PIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRT 1529
Db    685 -----------------LIDTEAEIDEVVSAFSAECER--------GETSGTKVLCKPLT 719

Qy   1530 CSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRAL 1589
Db    720 ------------------------------------------------------PPGF 723

Qy   1590 EDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRK 1649
Db    724 ENVLPAVK-------PLVSKGK----------------------TVKRV---------- 743

Qy   1650 GGKVMVINVASGDYFLMPCGF---QRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLA 1706
Db    744 ------------DYFQVMGGERLPKRPVVSGDNSVDARREFLYYLDAERVAQNDEIMSL- 790

Qy   1707 GGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG---REVKGRSTYSIWCEQDYV 1763
Db    791 -------------YRDYSRGVIRT-----GGQNYPHGLGVWDVEMKN------WCIRPVV 826

Qy   1764 RKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGI-ERTYRALLMARN 1822
Db    827 TEHAYV--------FQPD--------------KRMDDWSGYLEVAVWER---------- 853

Qy   1823 IAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIPHSTQDGLRVRDLPYVLIAEKGIFTKG 1882
Db    854 ------------------------------------GMLVNDF-------------- 860

Qy   1883 KDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGK 1942
Db    861 ----AVERMSDYVIVCDQTYLCNNRL-ILDNLSALDLGPVNCSF-----ELVDGVPGCGK 910

Qy   1943 TTTLVDEFVKSPNSTATITANVG-----------SSEDINMAVKKRDPNLEGLNSATTVN 1991
Db    911 STMIVN----SANPCVDVVLSTGRAATDDLIERFASKGFPCKLKRR---------VKTVD 957

Qy   1992 SRVVNFIVRGMYKRVL-VDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFR 2050
Db    958 SFLMHCVDGSLTGDVLHFDEALMAHAGMVYFCAQIAGAKRCICQGDQNQISFKPRVSQVD 1017

Qy   2051 MDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSK 2110
Db   1018 LRFSSLVGKFDIVTEKRETYRSPADVAAVLN--------KYYTGDVRT--HNATANSMTV 1067

Qy   2111 RPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRT 2170
```

```
Db   1068 RKIVSKEQVSLKPGAQYITFLQSEKKELVNLLALRKVAAKVSTVHESQGETFKDVVLVRT 1127

Qy   2171 KKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTF 2228
Db   1128 KPTDDSIARGREYLIVALSRHTQSLVYETVK---EDDVSKEIRESA--ALTKAALARF 1180


RESULT   7
WMFM12
125K protein - alfalfa mosaic virus
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (E
C;Species: alfalfa mosaic virus, AMV
C;Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 05-May-2000
C;Accession: A04197
R;Cornelissen, B.J.C.; Brederode, F.T.; Moormann, R.J.M.; Bol, J.F.
Nucleic Acids Res. 11, 1253-1265, 1983
A;Title: Complete nucleotide sequence of alfalfa mosaic virus RNA 1.
A;Reference number: A04197; MUID:83143345
A;Accession: A04197
A;Molecule type: mRNA
A;Residues: 1-1126 <COR>
A;Cross-references: GB:L00163; GB:J02000; NID:g331424; PIDN:AAA46289.1; PID:g331425
C;Genetics:
A;Map position: segment 1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine
F;99-188/Domain: methyltransferase #status predicted <MTF>
F;838-845/Region: nucleotide-binding motif A (P-loop)
F;845-1096/Domain: helicase #status predicted <HHG>


  Query Match             2.9%;  Score 329.5;  DB 1;  Length 1126;
  Best Local Similarity   16.9%;  Pred. No. 2.8e-12;
  Matches  318;  Conservative  191;  Mismatches  488;  Indels  887;  Gaps   63;

Qy    435 TKLCDMFSQ------RDAMIREKPSHRCDVF-----LKPREREKLRELFPELSIQFSDSV 483
Db     40 TAIGKVFSEAGRAYAQDAL----PSDKGEVLKISFSLDATQQNILRANFPGRRTVFSNSS 95

Qy    484 RSSHPFANAMRSCFNGIFSRRCGNV--CFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVK 541
Db     96 SSSHCFAAAHRLLETDFVYRCFGNTVDSIIDLGGNFVSHMKVKRHNVHCCCPILDARDGA 155

Qy    542 RRINEILFLSTAGGDSYVSS--DLLTEAASKSVSYCSRESQNCDSRADAGFMV-DVYDIS 598
Db    156 RLTERILSLK-----SYVRKHPEIVGEA-----DYCMDTFQKCSRRADYAFAIHSTSDLD 205

Qy    599 PQQVAEAMDKKGALVFDIALMFPVELLYGN-GEVYLEELDTLVKREGDYLAY------NV 651
Db    206 VGELACSLDQKGVMKFICTMMVDADMLIHNEGEIPNFNVRWEIDRKKDLIHFDFIDEPNL 265

Qy    652 GQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGT---E 708
Db    266 G-----YSHRFSLLKHYLTY--------NAVDLGHAAYR-------IERKQDFGGVMVID 305

Qy    709 VTYRSL-----VPSFVGKSLVFIPVVAGSSV-----------SFKTIVLDSDFVDRIYSY 752
Db    306 LTY-SLGFVPKMPHSNGRSCAWYNRVKGQMVVHTVNEGYYHHSYQTAVRRKVLVDKKVLT 364

Qy    753 ALNTIGTFENRTFEYAVGAVRSQKT-------HVITGSRVVHSKVDISPDDMWGLVVAVM 805
Db    365 RVTEVAFRQFRPNADAHSAIQSIATMLSSSTNHTIIGGVTLISGKPLSPDDYIPVATTIY 424

Qy    806 AQAIK------------DRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAK 853
Db    425 YRVKKLYNAIPEMLSLLDKGERLSTDAVLKGSEGPM--------WYSGPTFLSALD---K 473

Qy    854 AMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKN 913
Db    474 VNVPGDF-VAKALLSLPK-----------------------------------RDLKS 495

Qy    914 SRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKA 973
```

```
Db    496 ----LFSRS------------ATSHSERTPVRDE----------SPIRC----------- 518

Qy    974 TVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGI 1033
Db    519 ----------------------------TDGVFY---------------------PIRM 528

Qy   1034 LETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAA 1093
Db    529 L-------LKCLGSDK--FESVTITD---------------------------------- 545

Qy   1094 NARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVP 1153
Db    546 ----------------------PRSNTETTVD-----LYQSFQKKI------------- 564

Qy   1154 LALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAV 1213
Db    565 --------------------------ETVFSFILGKIDGP------------------- 578

Qy   1214 PEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLR--YYRDF 1271
Db    579 ------------------------------------SPLISDPVYFQSLEDVYYAEW 599

Qy   1272 DYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSV 1331
Db    600 --------------------HQGNAIDASNY---------------------------- 610

Qy   1332 QYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVD 1391
Db    611 ----------------------------------ARTLLDDIRKQKEESLKAKAKEVE 634

Qy   1392 GKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVH 1451
Db    635 DAQKLNRAIL-------------------------------------------------QVH 647

Qy   1452 AQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQR 1511
Db    648 AYLEAHPD----------------------GGKIEGLGLSSQFIAKIPELAIPTPKPL 683

Qy   1512 SLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGE 1571
Db    684 PEFEKNAETGEIL----------RINPHSDAIL-----EAIDYLKSTSANSIITLNKLG- 727

Qy   1572 GYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNP 1631
Db    728 ------------------------------DHC--QWTTKGLDVVWAGDDK-------- 746

Qy   1632 ILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFR 1691
Db    747 ------------------------------------------------------------ 746

Qy   1692 ATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGR 1751
Db    747 --RRAF-------------------IPKKNTWVGPTARSYPLAKYERAMS---------- 775

Qy   1752 STYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTG-----TSEDAVVEYLKYL 1806
Db    776 -------KDGYVT----LRWDGEVLDANCVRSLSQYEIVFVDQSCVFASAEAIIPSLE-- 822

Qy   1807 AIGIERTYRALLMARNIAVTTAEGVL---KVPN--QVYESLPGFHVYKSGTDLIFHSTQD 1861
Db    823 --------KALGLEAHFSVTIVDGVAGCGKTTNIKQIARS------------------- 854

Qy   1862 GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGM 1921
Db    855 -------------------SGRDVDLI--LTSNRSSADELKETIDCSPLTKLHYIRTC-- 891

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNL 1981
Db    892 --------------------------DSY------------------------------- 894

Qy   1982 EGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIP 2041
Db    895 --LMSASAVKA-----------QRLIFDECFLQHAGLVYAAATLAGCSEVIGFGDTEQIP 941
```

```
Qy   2042 FINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKD 2101
          |::|   |        | |  : |   :| | |  | |             ||
Db    942 FVSRNPSFVFRHHKLTGKVERKLIT---WRSPADATYCL------------EKYFYKNK 985

Qy   2102 KP------VVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVH 2155
          ||      |:||:   || :   |       :||| || ||: :|     || :  : | |
Db    986 KPVKTNSRVLRSIEVVPINSPVSVERNTNALYLCHTQAEKAVLKAQTHLKGCDN-IFTTH 1044

Qy   2156 EAQGKTFSDVVLFRTKKADDSLFTKQPHI------LVGLSRHTRSLVYAALSSKLDDKVG 2209
          ||||||| :|   |  :   || | :  |      || |||| ::  |  ::   || :
Db   1045 EAQGKTFDNVYFCRLTRTSTSLATGRDPINGPCNGLVALSRHKKTFKYFTIAHDSDDVIY 1104

Qy   2210 TYISDASPQSVSDALL-----HTF 2228
             ||   :  |::|     | |
Db   1105 NACRDAG--NTDDSILARSYNHNF 1126
```

RESULT 8
WMTM8T
180K protein - tomato mosaic virus (strain L)
N;Contains: 130K protein
C;Species: tomato mosaic virus
C;Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 30-Sep-1993
C;Accession: A04195
R;Ohno, T.; Aoyagi, M.; Yamanashi, Y.; Saito, H.; Ikawa, S.; Meshi, T.; Okada, Y.
J. Biochem. 96, 1915-1923, 1984
A;Title: Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and comp
A;Reference number: A91984; MUID:85157522
A;Accession: A04195
A;Molecule type: genomic RNA
A;Residues: 1-1615 <OHN>
A;Note: readthrough of the terminator UAG between codons for Gln-1116 and Gln-1117 occur
C;Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.7%; Score 309; DB 1; Length 1615;
Best Local Similarity 16.8%; Pred. No. 1.1e-10;
Matches 315; Conservative 205; Mismatches 519; Indels 840; Gaps 64;

```
Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
          |  |    |::: ::     | :  |     || : | |: ||     :|       :
Db      2 AYTQTATSSALLETVRGNNTLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVVSE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
          :    : :||  | | ::  : |  |  :||     :  :       :||||:|  |:
Db     57 EQTLIATKAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVS----------SDLLTE 566
          |    | | | |||:|: |   :     |:||    |: :|           ::: |
Db    117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGNKHVPNFQKEAFDRYAEMPNE 176

Qy    567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
              | || |   |  |  :  :|||   :   |: :|   |   |  |   ||
Db    177 VVCHDTFQTCRHSQECYTGRVYAIALHSIYDIPADEFGAALLRKNVHVCYAAFHFSENLL 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
            :  | |:|::   :|:|| | ::      : | ||:||:  :   :|   |:  |:
Db    237 LEDSHVNLDEINACFQRDGDRLTFSFASESTLNYSHSYSNILKYVCKTYFPASNREVYMK 296

Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
          |:
Db    297 EF------------------------------------------------------------ 298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
          | |: ::    : | ||                   :    | |  ||
Db    299 LVTRVNTWFCKFSRIDTF-------------------LLYKGVAHKGVD----------- 328

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
                    | | || | :       : :|:  | |  : :   : |:  |
Db    329 --------------SEQFYKAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
```

```
          :  :: :|                                  |||| ||     | :|  |
Db    369 MRDMVIVP--------------------------LFDISLET----SKRTRKEVL--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
          : :| :  |:  ::|   | |  |  :| :|:   | |      :: |  |     | |
Db    394 VSKDFVYTVLNHIRTYQAKAL--TYSNVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041
           |:   : |        ||                          ::: : :|:   : :
Db    446 SLLQSLSMT--------FF-----------------------LHTKLAVLKDD-LLI 470

Qy   1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097
          ||   | |      | |:||
Db    471 SKFALGPKTVSQHVWDEISL------------------------------------------ 490

Qy   1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157
                                                | :|    :  :  |:
Db    491 ------------------------------------AFGNAFPSIKERLINRK------ 507

Qy   1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE-- 1215
                  :: |  :||                                    ||:
Db    508 ------------LIKITENALE-----------------------------IRVPDLY 524

Qy   1216 VVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDE 1275
          | :  : ||   : : :                  | :: |   :||   |
Db    525 VTFHDRLVSEYKMSVDM-------------------PVLDIRKKMEETEEMY------- 557

Qy   1276 GAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTY 1335
                                  :::|   |: |    :   ::        :|:
Db    558 --------------------NALSELSVLKNSDKFDVDVFSQMC--------QSL---- 584

Qy   1336 VEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGL 1395
             || |  |             |  :  | | : |:   | | :    |  :
Db    585 ---EVDPMTA------------AKVIVAVMSNESGLTLTFEQPTEANVALAL------- 621

Qy   1396 LPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKE 1455
                    |:  :|:||                 : :|:|    : | :  |
Db    622 -----------QDSEKASDG---------------ALVVTSRDVEEPSIKGS------M 648

Qy   1456 VKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTI 1515
          : |: || :||      :  |  | |                  :  ::|  :|  :
Db    649 ARGELQLAGLSG------DVPESSYT-----------------RSEEIESLEQFHMAT 683

Qy   1516 NEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY 1575
               ||:|            ||       :  |   |:|:|    :
Db    684 ASSLIHKQMCSI----------VYTGPLKVQQMKN---FIDSLVASLSA---------- 719

Qy   1576 NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTV 1635
                    |: :::  :|  :  |  |  : | |
Db    720 -----------AVSNLVKILKDTAAID--LETRQKFG--------------------- 743

Qy   1636 NLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SIDEGR 1685
                              |::|||  : : |       |   ||    :      ||
Db    744 ------------------VLDVASKRWLVKPSAKNHAWGVVETHARKYHVALLEHDE-- 782

Qy   1686 ISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG 1745
                |||:
Db    783 -----------FGI---------------------------------------------- 785

Qy   1746 REVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805
                    | |:          | |                  | ::||    |
Db    786 ----------ITCD-------NWRR-------------------VAVSSESVV----Y 803

Qy   1806 LAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRV 1865
             :   || | ||      |::|: ||
Db    804 SDMAKLRTLRRLLKDGEPHVSSAKVVL------------------------------- 830

Qy   1866 RDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGES 1925
                         || |
Db    831 ------------------VDGV---------------------------------- 834

Qy   1926 FKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------VKKRD 1978
                  || ||              |  | : |   ||: :        : :|
```

```
Db    835 ------------PGCGK--------------TKEILSRVNFEEDLILVPGRQAAEMIRRR 868

Qy   1979 PNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
Db    869 ANASGIIVATKDNVRTVDSFLMNYGKGARCQFKRLFIDEGLMLHTGCVNFLVEMSLCDIA 928

Qy   2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTV 2085
Db    929 YVYGDTQQIPYINRVTGFPYPAHFAKLEVDEVETRRTTLRCPADVTHFLNQRYEGHVMCT 988

Qy   2086 RGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
Db    989 SSEKKSVSQEMVSGAASINPVSKPLKGK--------------ILTFTQ---SD-KEALL 1029

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYAA--- 2199
Db   1030 SRG-YADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYYTVVM 1088

Qy   2200 --LSSKLDD--KVGTYISD 2214
Db   1089 DPLVSIIRDLERVSSYLLD 1107


RESULT   9
S03701
141K protein - pea early browning virus
C;Species: pea early browning virus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C;Accession: S03701
R;MacFarlane, S.A.; Taylor, S.C.; King, D.I.; Hughes, G.; Davies, J.W.
Nucleic Acids Res. 17, 2245-2260, 1989
A;Title: Pea early browning virus RNA1 encodes four polypeptides including a putative zi
A;Reference number: S03701; MUID:89202038
A;Accession: S03701
A;Status: preliminary
A;Molecule type: genomic RNA
A;Residues: 1-1766 <MAC>
A;Cross-references: EMBL:X14006; NID:g61019; PIDN:CAB37343.1; PID:g4456716
C;Superfamily: cucumber mosaic virus RNA 1 protein


  Query Match             2.6%;  Score 304;  DB 2;  Length 1766;
  Best Local Similarity   17.2%;  Pred. No. 2.6e-10;
  Matches  331;  Conservative  237;  Mismatches  551;  Indels  802;  Gaps   75;

Qy    400 VANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDV- 458
Db      7 ISQLLNIDEFGAEQRGQFLDLMITKPDSQLGAMMQRILTDKI---DDSLRERKTRETVVI 63

Qy    459 --FLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNV-CFFDIGG 515
Db     64 HELLSQKDQNKLMEIYPEFNIVFKDDKNMVHGFAAAERKLQALLLMARVPKLEPVDDIGG 123

Qy    516 SFTYHVKAGHVNCHVCNPVLDVKDVKRRINEILFL-----STAGGDSYVSSDL--LTEAA 568
Db    124 QWSFWLSRGDKRVHSSCPLIDMRDKQRELQRQNFLRVFRDNATTSESRISDDQFDMYNAF 183

Qy    569 SKSVSY-----CSRESQNCDSRADAGFMVD-------------VYDISPQQVAEAMDKK 609
Db    184 KGDIDVANFVRCNNTFQDCNCR---GYRADGTRIGATNAIALHSLYDFKLDDVADAMIEK 240

Qy    610 GALVFDIALMFPVEL-------LYGNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSF 662
Db    241 GTKFLHAAMLFAPEAEIEKEGPLPSVDGYYERKEGSLISSEKIFFGFN-NDPSYAYIHDW 299

Qy    663 SNVSGFFTFSYVR----TSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVT----YRSL 714
Db    300 SEYK-----KYLRGEPFSRRGHVFMFEPWQARGDTMFFTLYRMTGVPMTNLLGNEYYRRL 354

Qy    715 -VPSFVGKSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVR 773
Db    355 HISRWEG--MVVVPVFEIDEIT-KKLTKSSMYVEKAYM----------DKCLDYV----- 396

Qy    774 SQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKS-IRSYNFIKASEGSLAGV 832
Db    397 ---------SRLSDQQLTI-----------------NNVKSFMSSNNWV---------- 419

Qy    833 FKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASD 892
Db    420 ----------LFINGAAVKNKQSVD-----------PR--------------------- 436

Qy    893 RLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAF 952
Db    437 RLQLLAQTLLVKE----------------------KLMRPLMTEMREKMVL-------- 465

Qy    953 IMGNVSNVHCTRAGLLGGSKATVVSSVSKGL-VARGAATKAFSGITSFFSTGSLFYDRGL 1011
Db    466 ------------------RATAVNSVTGTLKVAWG----------NYFD-GSL------ 489

Qy   1012 TEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLI 1071
Db    490 ----RRKCLLKLAQVLGTETGL----------EVLEVKDAPKYIEINDYLTAI------- 528

Qy   1072 GIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLV 1131
Db    529 --------------------------------FNETTEKTDGSLPDLDEA-------- 546

Qy   1132 VRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAIS------------------GCSMLG 1172
Db    547 --KRNSDKISKEAAEAAVQCVKMQFPKFESSDSLKEPLIRKIDGDVKKKGKQRRNCGLLT 604

Qy   1173 IWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRG-------KSVSS 1225
Db    605 GWTKLL----------------------------NNTHA--HVGWQRLFLKSKIESESE 633

Qy   1226 LLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHE 1285
Db    634 GRKPMT---DEEIEAALDDIM-----------ELNDVNLEAYK---------------- 662

Qy   1286 AVPGDDNDGSTSSVS-SYDVVT----NVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEV 1340
Db    663 ----------TRTVNKEFDIFTTWLASTYDTGLDSEKELI----------------TNL 695

Qy   1341 APSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESV 1400
Db    696 LATAAVRNKK------ALSDKLAMLIDVDDSVNSFLRSLSDTD---------------- 732

Qy   1401 VGEAPTQERGRAADGNT--AQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQ 1458
Db    733 -------------DDSTDVADCSASVSSDTSCVSSVVFRP--TVPPGFE-DVNLRK---- 772

Qy   1459 EVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEG 1518
Db    773 --------GKTVIVDNDAAESSSS--------------SERNR---------------- 793

Qy   1519 KAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGG 1578
Db    794 ---------------------------NHFAN-FEVIEN---------CRFGDAPKETGD 816

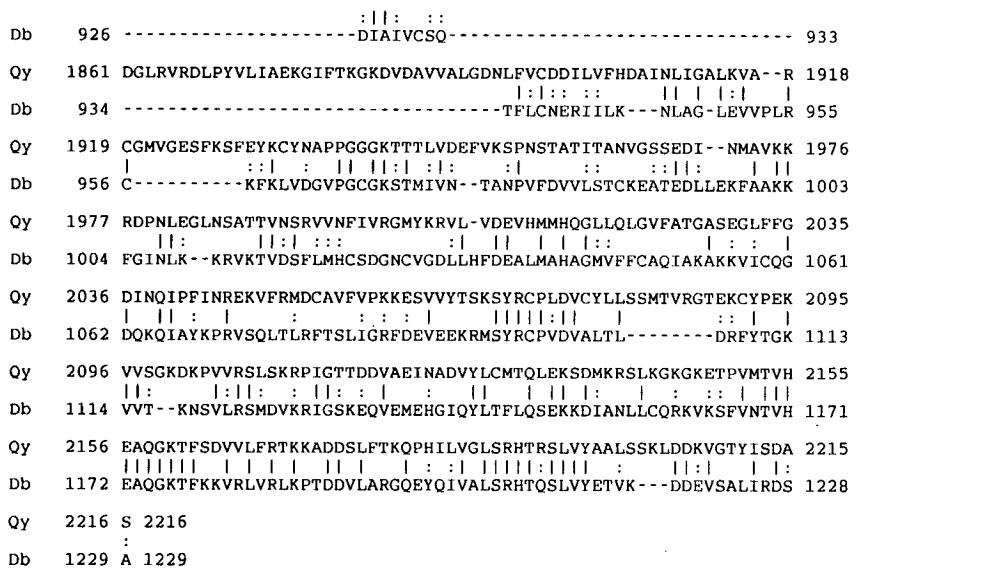
Qy   1579 SHVSSGWPRALEDILTAIKYPSVFDHCLV-----------QKYKMGGGVPFHADDEECYP 1627
Db    817 FSVDSR--------LEFIHYL----RCLICAQNNELLGKYRDYEMG------------- 850

Qy   1628 SDNPILTVNLVGKANFSTKCRKGGK-----VMVINVASGDYFLMP--CGFQRTHLHSVNS 1680
Db    851 ------------------VVRPGGKGYPDELGVFDLALKKWIIKPPSCSYNKAFVPDVSA 892

Qy   1681 IDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKAL 1740
Db    893 KEQGK------------WIGYLVDASWG-----------------------KQGIDAF 915

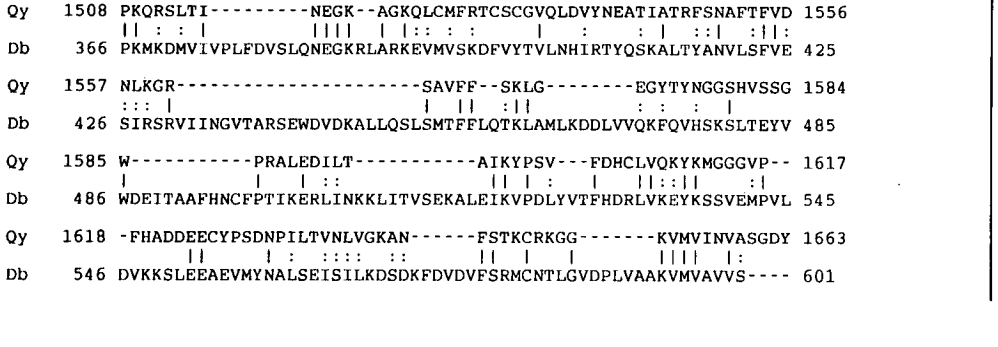
Qy   1741 SEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVV 1800
Db    916 S----------------------C---------------YTNVAWKA----------- 925

Qy   1801 EYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQ 1860
```



RESULT 10
WMTMPV
183K protein - pepper mild mottle virus (strain Spain)
N;Contains: 126K protein
C;Species: pepper mild mottle virus
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994
C;Accession: JQ1312
R;Alonso, E.; Garcia-Luque, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra, M.
J. Gen. Virol. 72, 2875-2884, 1991
A;Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistanc
A;Reference number: JQ1312; MUID:92113528
A;Accession: JQ1312
A;Molecule type: genomic RNA
A;Residues: 1-1611 <ALO>
A;Cross-references: GB:M81413
A;Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln and
C;Comment: This protein may have RNA polymerase activity.
C;Superfamily: cucumber mosaic virus RNA 1 protein
F;1-1117/Product: 126K protein #status predicted <PRO>

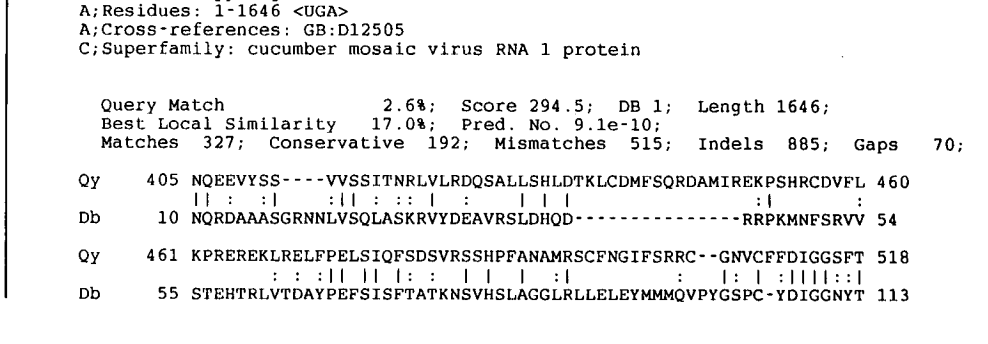
Query Match 2.6%; Score 297; DB 1; Length 1611;
Best Local Similarity 22.8%; Pred. No. 6.1e-10;
Matches 195; Conservative 114; Mismatches 301; Indels 244; Gaps 40;





RESULT 11
WMTMS2
186K protein - cucumber green mottle mosaic virus (strain SH)
C;Species: cucumber green mottle mosaic virus
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Apr-1994
C;Accession: JQ1157
R;Ugaki, M.; Tomiyama, M.; Kakutani, T.; Hidaka, S.; Kiguchi, T.; Nagata, R.; Sato, T
J. Gen. Virol. 72, 1487-1495, 1991
A;Title: The complete nucleotide sequence of cucumber green mottle mosaic virus (SH s
A;Reference number: JQ1157; MUID:91311400
A;Accession: JQ1157
A;Molecule type: genomic RNA
A;Residues: 1-1646 <UGA>
A;Cross-references: GB:D12505
C;Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.6%; Score 294.5; DB 1; Length 1646;
Best Local Similarity 17.0%; Pred. No. 9.1e-10;
Matches 327; Conservative 192; Mismatches 515; Indels 885; Gaps 70;



```
Qy    519 YHVKAGHVNCHVCNPVLDVKDVKRRI--NEILF--------------LSTAGGDSYVSSD 562
Db    114 QHLFKGRSYVHCCNPCLDLKDVARNVMYNDMITQHVQRHKGSCGCRPLPTFQIDAFRRYD 173

Qy    563 LLTEAASKSVSYCSRESQNCDSRADAG------FMVDVYDISPQQVAEAMDKKGALVFDI 616
Db    174 -----SSPCAVTCSDVFQECSYDFGSGRDNHAVSLHSIYDIPYSSIGPALHRKNVRVCYA 228

Qy    617 ALMFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVR 675
Db    229 AFHFSEALLLGSPVGNLNSIGAQFRVDGDDVHFLFSEESTLHYTHSLENIKLIVMRTYFP 288

Qy    676 TSSGNVFKLEYEGYRCGYHHLTMCRA-----QKSPGTEVTYRS------LVPSFVGKSL- 723
Db    289 ADDRFVYIKEFMVKRVDTFFFRLVRADTHMLHKSVGHYSKWKSEYFALNTPPIFQDKATF 348

Qy    724 ---------VFIPV-----VAGSSVSFKTIVLDSDFVDR1YSYALNTIGTFENR--TFEY 767
Db    349 SVWFPEAKKVLIPKFELSRFLSGNVKISRMLVDADFVHTI----INHISTYDNKALVWKN 404

Qy    768 AVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIR--SYNFIKAS 825
Db    405 VQSFVESIRSRVIVNGVSVKSEWNVPVDQLTDISFSIF-PLVKVRKVQIELMSDKVVIEA 463

Qy    826 EGSL---AGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVV 882
Db    464 RGLLRRFADSLKSAVEGLGDC------VY---------------------------- 486

Qy    883 VTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPL 942
Db    487 ---------DALVQTGWFDTS--------------------SDELKVL--------LPE 508

Qy    943 PITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFST 1002
Db    509 PFMTFSDYL------------EGMYEADAKIERESVSELLASGDDLFKKIDEIRNNYS- 554

Qy   1003 GSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTT 1062
Db    555 -------GVEFD-------------------------------VEKFQEFCKELNV----- 572

Qy   1063 FVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTS 1122
Db    573 ----NPMLIG-------------------------------HVIEAIFSQKAG------- 590

Qy   1123 STVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHL 1182
Db    591 ----------------VTVT---------------------------------------- 594

Qy   1183 MFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLL 1242
Db    595 ----GLGTL---------------------SPEM---GASVA----------------L 609

Qy   1243 NDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSY 1302
Db    610 SST---------------------------------------------SVDTCEDM 620

Qy   1303 DVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGSGTADAM 1362
Db    621 DVTEDMEDI-------VLMADKSHS-------YMSPEMARWA---------------- 648

Qy   1363 AFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERGRAADGNTAQTAV 1422
Db    649 ----DVKYG-----------------NNKGALVEYKVGTSMT---------------- 669

Qy   1423 NEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTT 1482
Db    670 -----------------LP-ATWAE-----KVKAVLPLSGICVRKP------------ 692

Qy   1483 RGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEA 1542
Db    693 --------------------QFSKP--------------------------LDEEDD- 703

Qy   1543 TIATRFSNA-FTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSV 1601
Db    704 ---LRLSNMNFFKVSDLK---------------------------LKKTITPVVYTGT 731

Qy   1602 FDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASG 1661
Db    732 IRERQMKNYI---------------------------------------------------- 741

Qy   1662 DYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQ 1721
Db    742 DYLSASLGSTLGNLERI---------------------------------------------- 758

Qy   1722 PQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIW-CEQDYVRKCEWLRADNPVMALEP 1780
Db    759 -----------------VRSDWNGTE-ESMQTFGLYDCE-----KCKWL--------LLP 787

Qy   1781 DYTPMTFEVVKTGTSEDAVVEYLKYLAIGI----ERTYRALLMARNIAVTTAEGVLKVPN 1836
Db    788 AEKKHAWAVVLASDDTTRII-FLSYDESGSPIIDKKNWKRFAVCSETKVYSVIRSLEVLN 846

Qy   1837 QVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLF 1896
Db    847 KEAIVDPGVHI----------TLVDGV---------------------------------- 863

Qy   1897 VCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLV------DEF 1950
Db    864 ----------------------------------------PGCGKTAEIIARVNWKTDL 882

Qy   1951 VKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIV--RGMYK--RV 2006
Db    883 VLTPGREAA------------AMIRRRACALHKSPVATNDNVRTFDSFVMNRKIFKFDAV 930

Qy   2007 LVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKK------ 2060
Db    931 YVDEGLMVHTGLLNFALKISGCKKAFVFGDAKQIPFINR--VMNFD----YPKELRTLIV 984

Qy   2061 ESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSK---------R 2111
Db    985 DNVERRYVTHRCPRDVTSFLNTI----------YKAAVATTSPVVHSVKAIKVSGAGILR 1034

Qy   2112 PIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTK 2171
Db    1035 P-----ELTKIKGKI-ITFTQSDKQSLIKS-----GYNDVNTVHEIQGETFEETAVVRAT 1083

Qy   2172 KADDSLFTK-QPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFA 2229
Db    1084 PTPIGLIARDSPHVLVALTRHTKAMVYYTV---VFDAVTSIIADV--EKVDQSILTMFA 1137
```

RESULT 12
JQ1896
hypothetical 191.1K protein - raspberry bushy dwarf virus
C;Species: raspberry bushy dwarf virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-2000
C;Accession: JQ1896
R;Ziegler, A.; Natsuaki, T.; Mayo, M.A.; Jolly, C.A.; Murant, A.F.
J. Gen. Virol. 73, 3213-3218, 1992
A;Title: The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus.
A;Reference number: JQ1896; MUID:93107856
A;Accession: JQ1896
A;Molecule type: mRNA
A;Residues: 1-1694 <ZIE>
A;Cross-references: GB:S51557; NID:g262149; PIDN:AAB24599.1; PID:g262150
C;Genetics:
A;Map position: segment RNA-1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: RNA replication

Query Match 2.5%; Score 291; DB 2; Length 1694;
Best Local Similarity 25.6%; Pred. No. 1.6e-09;
Matches 122; Conservative 65; Mismatches 186; Indels 104; Gaps 15;

```
Qy   1784 PMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLP 1843
Db    725 PETPEVTMTGNPNAVIAEAISYCRAEVDRIGK-----------KCERIL----HAYQATG 769

Qy   1844 GFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIF-------TKGKDVDAVVALGDNLF 1896
Db    770 NCGGYLNDTDNV--------------------GVFDKMTSWVQKPKEFDHEFGWDGSSF 808

Qy   1897 V-----------CDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTT 1945
Db    809 IKLSWFGKIPDFVGRYLVVTDGTRVTTNLKFSRQYATIPATVTPTIKLVDGVTGCGKTTE 868

Qy   1946 LVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKR 2005
Db    869 IVRRYRPGILILSVCKANV---DEIRRKLAAVDSKF-----IRTVDSYLLSPSVTGSCDE 920

Qy   2006 VLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESV-- 2063
Db    921 LFIDEYGLSHPGILLLAIHISGIRKVTLFGDSEQIPFCNR-------LADFPLKYNSVED 973

Qy   2064 -------VYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTT 2116
Db    974 VGLNFDREIRSTTYRCPQDITLSLQKM-----YKTKPIKTVS----TVESSITIKPIKSE 1024

Qy   2117 DDVAEINA----DVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKK 2172
Db   1025 FEIPLPNAFDGPVLYICMTKHDESLLKLRWAKENISSEVRTVHAAQGLSYKNVVYFRLTR 1084

Qy   2173 ADDSLFTKQ--PHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHT 2227
Db   1085 TDNDLYTKRKLPYHLVAISRHTDKIVYCT----------TKPEDSSDFSLS-ALKNT 1130
```

RESULT 13
P1BVBB
RNA 1a protein - broad bean mottle virus (strain Bawden)
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (E
C;Species: broad bean mottle virus
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 05-May-2000
C;Accession: A41699
R;Dzianott, A.M.; Bujarski, J.J.
Virology 185, 553-562, 1991
A;Title: The nucleotide sequence and genome organization of the RNA-1 segment in two brc
A;Reference number: A41699; MUID:92074218
A;Accession: A41699
A;Molecule type: genomic RNA
A;Residues: 1-966 <DZI>
A;Cross-references: GB:M65138; NID:g210662; PIDN:AAA42740.1; PID:g210663
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine
F;80-168/Domain: methyltransferase #status predicted <MTF>
F;688-944/Domain: helicase #status predicted <HHG>
F;690-697/Region: nucleotide-binding motif A (P-loop)

Query Match 2.5%; Score 286.5; DB 1; Length 966;
Best Local Similarity 30.9%; Pred. No. 1.1e-09;
Matches 105; Conservative 46; Mismatches 124; Indels 65; Gaps 13;

```
Qy   1920 GMVGESFKSFEY------KCYNAP----------------------PGGGKTTTLVDEFV 1951
Db    646 GGLGEKFENELYVVDKTCSCANAKAIADACKKVSAPTCSVVMVDGVAGCGKTTAIKETF- 704

Qy   1952 KSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVV----NFIVRGMYK--R 2005
Db    705 -RFEKDIIVTANRKSAEDVRKAI-------FGDASDSEVALKVVRTADSAIMHGLPECHR 756

Qy   2006 VLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMD-CAVFVPKKESVV 2064
Db    757 LLVDEAGLLHYGQLLAVADLCKCSEVLAFGDTEQISFKSRDATFRMKYCNIEYDKRDIV- 815

Qy   2065 YTSKSYRCPLDVCYLLSSMTVRGTEKC------YPEKVVSGKDKPVVRSLSKRPIGTTDD 2118
Db    816 --SKTFRCPQDV---VSAVKIL-KRKCANRSSKYNGWVSSSK---VEKSLSKSRIVSINQ 866

Qy   2119 VAEINADVYLCMTQLEKSDMKRSLKGKG-----KETPVMTVHEAQGKTFSDVVLFRTKKA 2173
Db    867 VSMEKHKFYLTMTEADKAALCSRAKDVGLDKTWVESNMETVHEAQGKAVDHVVLVRLKST 926

Qy   2174 DDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYIS 2213
Db    927 KCDLFKSEEYCLVALTRHKRTFEYLYNGDLGGDLISFYVT 966
```

RESULT 14
JA0074
RNA 1 protein - cucumber mosaic virus (strain Fny)
C;Species: cucumber mosaic virus, CMV
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 06-Feb-1998
C;Accession: JA0074
R;Rizzo, T.M.; Palukaitis, P.
J. Gen. Virol. 70, 1-11, 1989
A;Title: Nucleotide sequence and evolutionary relationships of cucumber mosaic virus
A;Reference number: JA0074; MUID:89279182
A;Accession: JA0074
A;Molecule type: genomic RNA
A;Residues: 1-993 <RIZ>
A;Experimental source: strain Fny
A;Note: the authors translated the codon CGU for residue 30 as Ala and AAU for residu
C;Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.4%; Score 277; DB 2; Length 993;
Best Local Similarity 22.8%; Pred. No. 4.8e-09;
Matches 150; Conservative 94; Mismatches 271; Indels 144; Gaps 25;

```
Qy   1566 FSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEEC 1625
Db    434 YEQLGKMYDMWNASSISK-WFAALTRPL------RVFFSSVV----------HALFPTL 475

Qy   1626 YPSDNPILTVNLVGKANFSTKCR-KGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEG 1684
Db    476 RPREEKEFLIKLSTFVTFNEECSFDGGEEWDV-ISSAAYV----------ATQAVTDG 522

Qy   1685 RISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGS 1744
Db    523 KILAAQKAEKL---AEKLAQPVSEVSD--SPETSSQTPDD---TADVC----------GR 564

Qy   1745 GREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLK 1804
Db    565 EREVSELDSLSAQTRSPITRVAE--RA------------------------TAMLEYAA 597

Qy   1805 YLAIGIERTYRALLMARNIAVTTAEGVLKVPNQV-----YESLPG-----FHVYKSGTDL 1854
Db    598 Y--------------EKQLHDTTVSNLKRIWNMAGGDDKRNSLEGNLKFVFDTYFTVDPM 643

Qy   1855 IFHSTQDGLRVRDLPYVLIAEKGIFTKG---KDVDAVVALGDNLFVCDDILVFHDAINLI 1911
Db    644 VNIHFSTGRWMRPVPEGIVYSVGYNERGLGPKSDGELYIVNSECVICNS-----ESLSTV 698

Qy   1912 GALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDIN 1971
Db    699 TRSLQAGTGTISQ---------VDGVAGCGKTTAIKSIF--EPSTDMIVTANKKSAQDVR 747

Qy   1972 MAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
Db    748 MALFKSSDSKEACTFVRTADSVLLNEC--PTVSRVLVDEVVLLHFGQLCAVMSKLKAVRA 805

Qy   2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTV----RG 2087
Db    806 ICFGDSEQIAFSSRDASFDMRFSKIIPDETSDADT--TFRSPQDVVPLVRLMATKALPKG 863

Qy   2088 TEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGK 2147
Db    864 THSKYTKWVSQSK---VKRSVTSRAIASVTLVDLDSSRFYITMTQADKA----SLNSRAK 916
```

```
Qy   2148 E---------TPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVY 2197
          |         : ||||:|| :   | | | |     || :  : || |:||  :  |
Db    917 EMNLPKTFWNERIKTVHESQGISEDHVTLVRLKSTKCDLFKQFSYCLVALTRHKVTFRY 975


RESULT  15
P1VXPJ
RNA 1 protein - peanut stunt virus (strain J)
N;Contains: ATP-dependent helicase (EC 3.6.1.-); mRNA (guanine-N7-)-methyltransferase (E
C;Species: peanut stunt virus
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 05-May-2000
C;Accession: JQ1451
R;Karasawa, A.; Nakaho, K.; Kakutani, T.; Minobe, Y.; Ehara, Y.
J. Gen. Virol. 73, 701-707, 1992
A;Title: Nucleotide sequence analyses of peanut stunt cucumovirus RNAs 1 and 2.
A;Reference number: JQ1451; MUID:92185483
A;Accession: JQ1451
A;Molecule type: genomic RNA
A;Residues: 1-1005 <KAR>
A;Cross-references: DDBJ:D11126; DDBJ:D01123; NID:g222417; PIDN:BAA01900.1; PID:g222418
C;Genetics:
A;Map position: segment 1
C;Superfamily: cucumber mosaic virus RNA 1 protein
C;Keywords: hydrolase; methyltransferase; P-loop; S-adenosylmethionine
F;81-174/Domain: methyltransferase #status predicted <MTF>
F;720-977/Domain: helicase #status predicted <HHG>
F;722-729/Region: nucleotide-binding motif A (P-loop)


  Query Match             2.4%;  Score 275.5;  DB 1;  Length 1005;
  Best Local Similarity   24.8%;  Pred. No. 6.1e-09;
  Matches  137;  Conservative   86;  Mismatches  229;  Indels  101;  Gaps   21;

Qy   1706 AGGVSD--------EKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIW 1757
          |  |||        ||:  : ||||  |  :: :|        :     :   ||  :|
Db    511 AWAVSDGKQLVAEKEKALELANQQP-VQNDSQPLT--------TVNLSTQDPPRSVTAI- 560

Qy   1758 CEQDYVRKCEWLRADNPVMAL-----EPDYTPMTFEV----VKTGTSEDAVVEYLKYLAI 1808
                    |  :|||:        |:   ::|:      :      |::||: |
Db    561 -----------LDPENPVVGQAELPEPPELKALSFQTRSPDTRLAERASAMLEYVAY--- 606

Qy   1809 GIERTYRALLMARNIAVTTAEGVLKVPNQVYESLP-GFHVYKSGTDLIFHSTQDGLRVRD 1867
            |:     :::    :  : |   | |::  :|   |  |    |:       |  :|
Db    607 --EQQLHQNVISNLQRIWCSAGGDFVTNRLEGNLKFVFDTYFLTDPLVNVHFPSGRWMRP 664

Qy   1868 LP-----YVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALK--VARCG 1920
          :|      |   |||:  |         :   |:: :   |  ::  |  |:|   |  |
Db    665 VPAGTKYMVGFNEKGLGPK---------MDSELYIVNADCVICNSDGLSSAVKALTAPTG 715

Qy   1921 MVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPN 1980
           :           :   | |||| :   |   | :   :|||  |::|: ||:     :
Db    716 TI---------SLVDGVAGCGKTTAIKKLF--QPATDLIVTANKKSAQDVRMALFGASDS 764

Qy   1981 LEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQI 2040
           |       | :| ::| :       |||||| ::| | |    :     |   : |||  ||
Db    765 KEACTFVRTADSILLNDVPE--VSRVLVDEVVLLHFGQLCAVMSKLKAVRAICFGDSEQI 822

Qy   2041 PF1NREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMT----VRGTEKCYPEKV 2096
           | :|:  | |  :  :| : |   :  ::| | ||  |:: |      || :  |   :
Db    823 AFGSRDASFDMRRSTILPDESST--SDTTFRSPQDVVPLVTRMAGKALPRGAKSKYSRWL 880

Qy   2097 VSGKDKPVVRSLSKRPIGTTDDVAEINAD---VYLCMTQLEKSDMKRSLKGKGKE----- 2148
             |   |: |:|:| |     || :: |    |: ||| :|:    :|  : ||
Db    881 SQSK---VLNSVSQRCI---PSVALVDMDESRFYITMTQADKA----ALITRAKELQMPS 930

Qy   2149 ----TPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKL 2204
               : ||||:|| :   | | | |     ||   |: || |:||  |  |
Db    931 SFYDNKIKTVHESQGISEDHVTLVRLKSTKCDLFKSFPYCLVALTRHKVSFRYEFCGELK 990

Qy   2205 DDKVGTYISDASP 2217
           | :   ::  ||
Db    991 GDLIALVLATHSP 1003
```

Search completed: January 31, 2001, 15:28:52
Job time: 181 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:33:10 ; Search time 30.65 Seconds
(without alignments)
2356.995 Million cell updates/sec

Title: US-09-301-906-15
Perfect score: 11520
Sequence: 1 MDYIRPLRVFSFPHVNNTLE.....QSVSDALLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	402.5	3.5	1547	1	RRPO_PMV	P20951 papaya mosa
2	400	3.5	1718	1	RRPO_SHVX	Q04575 shallot vir
3	338.5	2.9	1707	1	194K_TRVSY	P05080 tobacco rat
4	337	2.9	1629	1	RRPO_SHMV	P89202 sunn-hemp m
5	329.5	2.9	1126	1	V125_AMVLE	P03589 alfalfa mos
6	313	2.7	1616	1	RRPO_TOMS1	Q9ykd6 tomato mosa
7	309	2.7	1616	1	RRPO_TOML	P03587 tomato mosa
8	303	2.6	1616	1	RRPO_TOMK2	P89676 tomato mosa
9	302	2.6	1616	1	RRPO_TOMK1	Q9q1t8 tomato mosa
10	298	2.6	1648	1	RRPO_CGMVS	P19523 cucumber gr
11	297	2.6	1612	1	RRPO_PMMVJ	P89657 pepper mild
12	297	2.6	1612	1	RRPO_PMMVS	P29098 pepper mild
13	288	2.5	1616	1	RRPO_TMVRA	Q98745 tobacco mos
14	286.5	2.5	966	1	V1A_BBMV	Q00020 broad bean
15	285	2.5	1616	1	RRPO_TMVB	Q9yq08 tobacco mos
16	279	2.4	1612	1	RRPO_ORSVS	Q84133 odontogloss
17	277	2.4	1612	1	RRPO_ORSVC	P89659 odontogloss
18	275.5	2.4	1005	1	V1A_PSVJ	P28726 peanut stun
19	275	2.4	1616	1	RRPO_TMVOM	O93058 tobacco mos
20	274	2.4	993	1	V1A_CMVFN	P17769 cucumber mo
21	273	2.4	993	1	V1A_CMVNT	O40976 cucumber mo
22	273	2.4	993	1	V1A_CMVO	P20122 cucumber mo
23	272.5	2.4	1609	1	RRPO_TMGMV	P18339 tobacco mil
24	271.5	2.4	1139	1	VRNA_BSMV	P17595 barley stri
25	270.5	2.3	961	1	V1A_BMV	P03588 brome mosai
26	270.5	2.3	1616	1	RRPO_TMOB	P90211 tobamovirus
27	270	2.3	993	1	V1A_CMVII	Q83270 cucumber mo
28	269	2.3	993	1	V1A_CMVIX	Q66121 cucumber mo
29	267	2.3	1616	1	RRPO_TMV	P03586 tobacco mos
30	264	2.3	1616	1	RRPO_TMVKR	P30738 tobacco mos
31	263	2.3	991	1	V1A_CMVQ	P06011 cucumber mo
32	259.5	2.3	1885	1	RRPO_ACLSA	P54891 apple chlor
33	253.5	2.2	993	1	V1A_CMVY	Q83264 cucumber mo
34	252	2.2	958	1	V1A_CCMV	P27752 cowpea chlo
35	245.5	2.1	1884	1	RRPO_ACLSP	P27738 apple chlor
36	238	2.1	1601	1	RRPO_TVCV	Q88920 turnip vein
37	237	2.1	993	1	V1A_TAV	P28931 tomato aspe
38	218	1.9	1597	1	RRPO_CRMV	Q66220 chinese rap
39	184.5	1.6	1693	1	POLN_HEVMY	Q04610 hepatitis e
40	179.5	1.6	1693	1	POLN_HEVPA	P33424 hepatitis e
41	173.5	1.5	1848	1	CBPA_CLOCL	P38058 clostridium
42	171	1.5	1691	1	POLN_HEVME	Q03495 hepatitis e
43	169.5	1.5	1693	1	POLN_HEVBU	P29324 hepatitis e
44	164.5	1.4	1608	1	HLYA_SERMA	P15320 serratia ma
45	163	1.4	6359	1	BACC_BACLI	O68008 b bacitraci

ALIGNMENTS

RESULT 1
RRPO_PMV
ID RRPO_PMV STANDARD; PRT; 1547 AA.
AC P20951;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICATION PROTEIN (176 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
DE DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
OS Papaya mosaic potexvirus (PMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89381685; PubMed=2778435;
RA Sit T.L., Abouhaidar M.G., Holy S.;
RT "Nucleotide sequence of papaya mosaic virus RNA.";
RL J. Gen. Virol. 70:2325-2331(1989).
CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; D13957; BAA03050.1; -.
DR PIR; JQ0096; JQ0096.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicase1; 1.
KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW Transferase.
FT NP_BIND 822 829 ATP (POTENTIAL).
SQ SEQUENCE 1547 AA; 176320 MW; 6CA4282C6A082622 CRC64;

Query Match 3.5%; Score 402.5; DB 1; Length 1547;
Best Local Similarity 24.0%; Pred. No. 1.8e-16;
Matches 177; Conservative 87; Mismatches 230; Indels 243; Gaps 33;

```
Qy   1432 SSLVSSPQADIPKVTQSEVHAQK-EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDK 1490
          | |   |:    |  ::    |: |||:|   |: ||    : | ||           |
Db    466 SLLQKEPEVMEAKEAEATDEPQRPEVKEEQAEASTSGRAEEIQEDPATK----------K 515

Qy   1491 GKAVAHVAEKKQVQVEQPK-QRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFS 1549
          ||             |:|   | |                    | ||: | : |      |
Db    516 GK------------EEPNPNRDL--------------LCPCGLHLKIKN-----AEFP 542

Qy   1550 N--AFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTA--IKYP-SVFDH 1604
               |:| || | |||| |:|:| |||| | |||  || || |  || |   |:
Db    543 ELPVLDHPDHLTGRKAWFFSKDGKPYSYTGGSHASRGWPNWLEKILAAIEIKEPLPEFNQ 602

Qy   1605 CLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYF 1664
          ||||::|:   :||| ||| |||  : :||:|  |:      |:||     | :  |||:
```

```
Db    603 CLVQQFKLQAAIPFHRDDEPCYPKGHQVLTINHSGECLTQIACQKG--KASITMGFGDYY 660

Qy   1665 LMPCGFQRTHLHSVNSIDEGRISLTFRAT--------------------------RRVF 1697
          | | ||| :| |:|::   ||:|||| |                          : :
Db    661 LSPVGFQESHKHAVSNTTGGRVSLTFRCTVQQNKFNDDGSMEALDNLPWKAWIPKLQNLG 720

Qy   1698 GVGRMLQ------LAGGVSDEKSPGVPNQQPQS---------QGATRTITPKSGGKALSE 1742
            || ||      |   : : :|  :|  :|:             | |  || |     :
Db    721 FQGRQLQYDPNGALISPIEEIRS--MPKCKPEGVPEVVYKTLDGLARAPTPYSPNPIRAR 778

Qy   1743 GSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEY 1802
             :||     ::  :|      ||                          |   ||:||
Db    779 AYTSDVKNCRIGALLRQQGK----EW----------------------GCRFDALVEA 810

Qy   1803 LKYLAIGIERTYRALLMARNIAVTTAEGV-LKVPNQVYESLPGFHVYKSGTDLIFHSTQD 1861
           |                | :|::   |     :|  ::|
Db    811 GK---------------RELAISVIHGAGGSGKSQALQTL------------------- 835

Qy   1862 GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGM 1921
             ::| |              ::|  | |  |             : |    |: :
Db    836 ---IKDNP-------------ELDITVVLPTN-----------ELRLDWLRKLPKAPQ 866

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPN----------STATITANVGSSEDIN 1971
            | ||:|| |   |||       : |:: | |           ||  :    |
Db    867 --EKFKTFE-KALLAPP---TPIVIFDDYGKLPAGYVEAFCLYFSTVQLIILTGDC---- 916

Qy   1972 MAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYK--RVLVDEVHMMHQGLL-QLGVFA--T 2026
            |:  : |   :||| :   :  :|:   :  |  ::  |   : |  :|||::  |
Db    917 ----KQSVHHESNENATTSS---IEPLVKEASELCRYYINATHRNKKDLANKLGVYSEKT 969

Qy   2027 GASE---------GLFFGDINQIPFINREKVFR---------MDCAVFVPKKESVVYTSK 2068
            | :|         ||       :| : ::: |           |      |  :: | :
Db    970 GLTEVTHGTTPIPGLHM----LVPSLYKKQAFSEMGHKVSTYAGCQGITAPKIQILLTEE 1025

Qy   2069 SYRCPLDVCYLLSSMTV 2085
          :  |  :| |   |  |
Db   1026 TSLCSREVLYTALSRAV 1042


RESULT   2
RRPO_SHVX
ID   RRPO_SHVX      STANDARD;      PRT;  1718 AA.
AC   Q04575;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA REPLICATION PROTEIN (194 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
DE   DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
OS   Shallot virus X (ShVX).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93019008; PubMed=1339468;
RA   Kanyuka K.V., Vishnichenko V.K., Levay K.E., Kondrikov D.Y.,
RA   Ryabov E.V., Zavriev S.K.;
RT   "Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal
RT   cistron closely related to those of potexviruses and a unique
RT   arrangement of the 3'-proximal cistrons.";
RL   J. Gen. Virol. 73:2553-2560(1992).
CC   -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
CC       POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M97264; AAA47787.1; -.
DR   PIR; JQ1734; JQ1734.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
KW   Transferase.
FT   NP_BIND     915    922       ATP (POTENTIAL).
SQ   SEQUENCE   1718 AA;  194529 MW;  820FEAE1EB62415C CRC64;


  Query Match             3.5%;  Score 400;  DB 1;  Length 1718;
  Best Local Similarity   23.5%;  Pred. No. 3e-16;
  Matches  187;  Conservative  114;  Mismatches  290;  Indels  206;  Gaps   37;

Qy   1530 CSCGVQLDVYNEATIATRFSNAF-----TFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSG 1584
          | || :: |         |  |         |::||| | |:|: |:||:| | || | |
Db    583 CECGTEITV-------NSFGRAIEVAGVNLTDHMKGRLAAFYSRDGQGYSYTGYSHKSQG 635

Qy   1585 WPRALEDILTAI-KYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANF 1643
          |   |: :: |  : |: :: |||||: |  : ||:|::  ||  | |||||  |   |
Db    636 WLEGLDKLIEACGEKPTTYNQCLVQKYEQGSRIGFHSDEQAIYPKGNKILTVNAAGSGTF 695

Qy   1644 STKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRML 1703
            || ||     :|:  |||| || ||| || |:| ::   |:| |||:|
Db    696 GIKCAKG--ETTLNLEDGDYFQMPSGFQETHKHNVVAVTP-RLSFTFRST---------- 742

Qy   1704 QLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGRE------VKGRSTYSIW 1757
                |: :| |  | :  |:    :  |       |  ||::       ||
Db    743 ----VVNSQKKPAEPEKLNQNNACPKPSDP-------SNASGKQHKKTHPAKGNE----- 786

Qy   1758 CEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVK----------TGTSEDAVVEYLKYL 1806
                      :: :|   |||   | | |::|           |  :   | : |
Db    787 -----------KSSSP--NLEPLDAP-TVEILKLHGFTALTPQHDGTCQIRPVYFNK-- 829

Qy   1807 AIGIERTYRALLM---AR---NIAVTTAEGVL--KVPNQ--------VYESLPGFHVYKS 1850
          | : |      |   ||   ::| :   |:   |: |:         |  :| | : |
Db    830 DIHLRRKAVKTDMSPPARPFFDLATSLHRGIYTHKIDNRRATAYMSDVKNNLTGLVLPKL 889

Qy   1851 GTDLIFHSTQDGL---RVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDA 1907
            ||:  |:   |     |::  : |   |   | : :  ::      |   | | :
Db    890 DRDLL--SSWVALAETTTREVAVLAIHGAGGAGKSRALQELLRSSPEL--ADSINIVVPT 945

Qy   1908 INLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSS 1967
          |||    |     |      :|:  |         |: |: |:: | |     : | :
Db    946 INLANDWKAKLPQMDPRRVMTFQKACERE----CKSVTIFDDYGKLP--AGFVDAYLAIK 999

Qy   1968 EDINMAV----------KKRDPNLEGLNSATTVNSRVVNFIVRGMY------------- 2003
           :: :|:           ::|:  :  | |     |:  :: :   :
Db   1000 VNVELAILTGDQRQSTHHQERESQISSLQSNIAQFSKYADYYLNATHRQPRRLANPIKVH 1059

Qy   2004 ----------KRVLVDEVHMM------HQGLL-QLGVFAT--GASEGLFFGDINQI---- 2040
                    |  :| :: |:       | ||  ||  |      :||    :  |
Db   1060 AERQLGGAVLKANIVPDLAMVLVPAFRSQSLLTDLGRHAMTYAGCQGLTLNHLTIILDKD 1119

Qy   2041 -PFINREKVFRMDCAVFVPKKESVVYT-----SKSYRCPLDVCYLLSSMT--VRGTEKCY 2092
           |  : | ::      |    ||: :      : ::   ||    | ::   ||  |:
Db   1120 TPLCSDEVLY----TAFSRASESITFVNTHSDNPAFLAKLDATPYLKTLISWVREDEE-- 1173

Qy   2093 PEKVVSGKDKPVVRSLSKR-PIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPV 2151
                :| | |    | |  |  |   ||      :   :  :|  | :    |: |   :
Db   1174 -----AGADCPATEPLVKDVPTKTHIPVANDKVQLEGKIEAMEDKDTRELWSGEEKTNLM 1228

Qy   2152 MTVHEAQGKTFSDVVLFRTKKA-DDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGT 2210
           |          | ||  ::| |::||                          |:   :|
Db   1229 QTQDPV-------VQLFPHQQAKDEALF-----------------------KI--TIGE 1255

Qy   2211 YISDASPQSVSDALLHT 2227
           |  |:|:  :  | ||
Db   1256 RIRMATPEQNAKQLRHT 1272


RESULT   3
194K_TRVSY
ID   194K_TRVSY      STANDARD;      PRT;  1707 AA.
AC   P05080;
DT   13-AUG-1987 (Rel. 05, Created)
```

```
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   01-JUL-1989 (Rel. 11, Last annotation update)
DE   POTENTIAL 194 KDA PROTEIN (PUTATIVE REPLICASE) [CONTAINS: 134 KDA
DE   PROTEIN].
OS   Tobacco rattle virus (strain SYM).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88034943; PubMed=3668507;
RA   Hamilton W.D.O., Boccara M., Robinson D.J., Baulcombe D.C.;
RT   "The complete nucleotide sequence of tobacco rattle virus RNA-1.";
RL   J. Gen. Virol. 68:2563-2575(1987).
RN   [2]
RP   PARTIAL SEQUENCE FROM N.A.
RA   Boccara M., Hamilton W.D.O., Baulcombe D.C.;
RT   "The organisation and interviral homologies of genes at the 3' end of
RT   tobacco rattle virus RNA1.";
RL   EMBO J. 5:223-229(1986).
CC   -!- SIMILARITY: THIS PROTEIN IS HOMOLOG TO PUTATIVE REPLICASE OF
CC       TOBACCO MOSAIC VIRUS (TMV) AND TRICORNAVIRUSES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06172; CAA29537.1; -.
DR   EMBL; D00155; BAA00110.1; -.
DR   PIR; S01865; S01865.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
FT   CHAIN         1   1187       134 KDA PROTEIN.
FT   CHAIN         1   1707       194 KDA PROTEIN.
SQ   SEQUENCE   1707 AA;  194328 MW;  F4DABDB6C7F51F04 CRC64;
```

```
  Query Match             2.9%;  Score 338.5;  DB 1;  Length 1707;
  Best Local Similarity   18.0%;  Pred. No. 1.9e-12;
  Matches  334;  Conservative  209;  Mismatches  514;  Indels  801;  Gaps   72;

Qy    445 DAMIREKPS------HRCDVFLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFN 498
Db     50 DDMIRERKTKDPVIVHE---VLSQKEQNKLMEIYPEFNIVFKDDKNMVHGFAAAERKLQA 106

Qy    499 GIFSRRCGNVC-FFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDS 557
Db    107 LLLLDRVPALQEVDDIGGQWSFWVTRGEKRIHSCCPNLDIRDDQREISRQIFLTAIGDQA 166

Qy    558 YVSSDLLTE----------AASKSVSYCSRESQNCDSRADAGF----------MV 592
Db    167 RSGKRQMSENELWMYDQFRKNIAAPNAVRCNNTYQGCTCR---GFSDGKKKGAQYAIALH 223

Qy    593 DVYDISPQQVAEAMDKKGALVFDIALMF-PVELLYGNGEVYLEELDTLVKREGDYLAYNV 651
Db    224 SLYDFKLKDLMATMVEKKTKVVHAAMLFAPESMLVDEG--------PLPSVDGYYMKKN- 274

Qy    652 GQCGEMYEHSFSNVSGFF---TFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTE 708
Db    275 ---GKIY-------FGFEKDPSFSYIHDWE------EYKKYLL----------GKP 304

Qy    709 VTYRSLVPSFVGKSLVFIP-VVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEY 767
Db    305 VSYQ-------GNVFYFEPWQVRGDTMLFS----------IYRIA---------- 332

Qy    768 AVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEG 827
```

```
Db    333 --GVPR----------RSLSSQEYYR---- 346

Qy    828 SLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICT 887
Db    347 ----------RIYISRWENMVVVPI-- 361

Qy    888 SGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITED 947
Db    362 ----------FDLVEST----------RELVKKDLFVEKQFM----------D 384

Qy    948 SVYAFIMG------NVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFS 1001
Db    385 KCLDYIARLSDQQLTISNV----------KSYLSSNNWVLFINGAAVK---------- 422

Qy   1002 TGSLFYDRGLTEDERLDALVRT---ENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLN 1058
Db    423 ------NKQSVDSRDLQLLAQTLLVKEQVARPV-----MRELREAILTETKPI---TSLT 468

Qy   1059 DFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLR 1118
Db    469 D----------VLGLISRKL------WK----------QFA----------NKIAVGGFV 492

Qy   1119 GLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHAL 1178
Db    493 GMV------GTLI---GFYPKKVLT----------WAKDT 513

Qy   1179 PRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITL-----G 1233
Db    514 PNG----------PELCYENSHKTKVIVFLSVVYAIGG 541

Qy   1234 VSLIIRGLLNDTIPQLAYVPPVE-GRNVYD--ETLRYYRDFDYDEGAGPSGTQHEAVPGD 1290
Db    542 ITLMRRDIRDGLVKKLCDMFDIKRGAHVLDVENPCRYY----------E 580

Qy   1291 DNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQ 1350
Db    581 INDFFSSLYSA----------SESGET----------VLPDLSEVKAK 608

Qy   1351 GDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERG 1410
Db    609 SD----------KLLQQKKEIADEFLSAKFSNYSGSSVRTS-----PPSVVGSSRSGLGL 653

Qy   1411 RAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEV-KQEVPLATVSGAT 1469
Db    654 LLEDSNV----------LTQARVGVSRKVDDEEIMEQFLSG-- 684

Qy   1470 PIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRT 1529
Db    685 ----------LIDTEAEIDEVVSAFSAECER--------GETSGTKVLCKPLT 719

Qy   1530 CSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRAL 1589
Db    720 ----------PPGF 723

Qy   1590 EDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRK 1649
Db    724 ENVLPAVK-------PLVSKGK----------TVKRV---------- 743

Qy   1650 GGKVMVINVASGDYFLMPCGF---QRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLA 1706
Db    744 ----------DYFQVMGGERLPKRPVVSGDNSVDARREFLYYLDAERVAQNDEIMSL- 790

Qy   1707 GGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG---REVKGRSTYSIWCEQDYV 1763
Db    791 ----------YRDYSRGVIRT-----GGQNYPHGLGVWDVEMKN------WCIRPVV 826

Qy   1764 RKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGI-ERTYRALLMARN 1822
Db    827 TEHAYV--------FQPD----------KRMDDWSGYLEVAVWER---------- 853

Qy   1823 IAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKG 1882
Db    854 ----------GMLVNDF---------- 860
```

```
Qy   1883 KDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGK 1942
Db    861 ----AVERMSDYVIVCDQTYLCNNRL-ILDNLSALDLGPVNCSF-----ELVDGVPGCGK 910

Qy   1943 TTTLVDEFVKSPNSTATITANVG----------SSEDINMAVKKRDPNLEGLNSATTVN 1991
Db    911 STMIVN----SANPCVDVVLSTGRAATDDLIERFASKGFPCKLKRR---------VKTVD 957

Qy   1992 SRVVNFIVRGMYKRVL-VDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFR 2050
Db    958 SFLMHCVDGSLTGDVLHFDEALMAHAGMVYFCAQIAGAKRCICQGDQNQISFKPRVSQVD 1017

Qy   2051 MDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSK 2110
Db   1018 LRFSSLVGKFDIVTEKRETYRSPADVAAVLN--------KYYTGDVRT--HNATANSMTV 1067

Qy   2111 RPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRT 2170
Db   1068 RKIVSKEQVSLKPGAQYITFLQSEKKELVNLLALRKVAAKVSTVHESQGETFKDVVLVRT 1127

Qy   2171 KKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTF 2228
Db   1128 KPTDDSIARGREYLIVALSRHTQSLVYETVK---EDDVSKEIRESA--ALTKAALARF 1180
```

```
RESULT   4
RRPO_SHMV
ID   RRPO_SHMV      STANDARD;      PRT;  1629 AA.
AC   P89202; Q88624;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (128 KDA PROTEIN)].
OS   Sunn-hemp mosaic virus (SHMV) (TMV strain cowpea).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE OF 1-1536 FROM N.A.
RX   MEDLINE=97093421; PubMed=8938983;
RA   Silver S., Quan S., Deom C.M.;
RT   "Completion of the nucleotide sequence of sunn-hemp mosaic virus: a
RT   tobamovirus pathogenic to legumes.";
RL   Virus Genes 13:83-85(1996).
RN   [2]
RP   SEQUENCE OF 1538-1629 FROM N.A.
RX   MEDLINE=83064542; PubMed=6292867;
RA   Meshi T., Ohno T., Okada Y.;
RT   "Nucleotide sequence of the 30K protein cistron of cowpea strain of
RT   tobacco mosaic virus.";
RL   Nucleic Acids Res. 10:6111-6117(1982).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR LYS-1130 AND GLN-1132.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U47034; AAB38492.1; -.
DR   EMBL; J02413; AAA46585.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1629       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1130       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     849    856       ATP (POTENTIAL).
SQ   SEQUENCE   1629 AA;  186513 MW;  C928BC9AF95FE8A5 CRC64;
```

```
  Query Match             2.9%;  Score 337;  DB 1;  Length 1629;
  Best Local Similarity   18.3%;  Pred. No. 2.2e-12;
  Matches  348;  Conservative  177;  Mismatches  505;  Indels  868;  Gaps   73;

Qy    427 SALLSHLDTKLCDMFSQRDAMIRE-------------KPSHRCDVFLKPREREKLRELF 472
Db      4 STLINKAQTNSCGDVGVVDLLKRKVYDDTVKTMQGLDRRAKYRLNQCLGPEQCRTVRGGY 63

Qy    473 PELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHVKAGHVNCHVC 531
Db     64 PEFQIEFTGASNTSHAMAAGLRGLELEYLYTLVPYGAVSYDIGGNFPAHMMKGRSYVHCC 123

Qy    532 NPVLDVKDVKRRINEILFLSTAGGDSYVSSDLLTEAASKSVSYCS-------------RE 578
Db    124 NPALDARDLARNENYRISI----------ENYLSRFEDKSGDYCQWQRKKPKVSKPLPRY 173

Qy    579 SQNCDSRAD--------------------------AGFMVDVYDISPQQVAEAMDKKGA 611
Db    174 QKACFDRYNEDPEHVTCSETFEKCRISPPAERDDIYATSLHSLYDIPYQNLGPALARKRI 233

Qy    612 LVFDIALMFPVELLYGNGEVYLEELDTLVKREGDYLAYN-VGQCGEMYEHSFSNVSGFFT 670
Db    234 KVLHAAFHFSEDLLLGASEGLLTQIGGTFQRNGDVLTFSFLDESSLIYTHSFRNVFEYVT 293

Qy    671 FSYVRTSSGNVFKLEYEGYR-----CGYHHL-TMCRAQKSPGTEVTYRSLVPSFVGKSLV 724
Db    294 RTFFVACNRYAYMKEFRSRRVDTVFCSFIRIDTYC----------LYRSV---------- 333

Qy    725 FIPVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSR 784
Db    334 -----------FK----DCD--------------------EHVFAAM----------- 345

Qy    785 VVHSKVDISPDDMWGL-VVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDC 843
Db    346 ----------DDAWEFKKKRVMLEASRPIFNDVAQFNVY-------------------- 374

Qy    844 FSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDIS 903
Db    375 FPNA---------------KDKVCLPIFAVKSVSGAPVTT--------RHIL-----VE 405

Qy    904 KETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCT 963
Db    406 KDFYWTALNH-----------------ILTYPDGK---------------------- 423

Qy    964 RAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRT 1023
Db    424 --------------------------ADFRGVMSFLES-----------------IRS 438

Qy   1024 ENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPI 1083
Db    439 RVVIN---GTTTASQWEVDK-----------------SQLKDIALSLLLIAKLEKLKI 476

Qy   1084 AWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVT 1143
Db    477 SVIEKR---------IKIERQGLVSLLKEFLHGL--LDEYTQTMAEWVVEKGWVKSVD-- 523

Qy   1144 RATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKF 1203
Db    524 --QVLQVTIPDLVLNFRDHFR---CEF------------------------RTSAN---- 550

Qy   1204 GGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDE 1263
Db    551 ---------VSEV-----NVSEHLVA--------------TNEYYAKVSDLVDRN---P 578

Qy   1264 TLRYYRDF----DYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNG-- 1317
```

```
Db    579 TLAF--DFEKFQDYCEKLG--------------------VDIDTVTELID-AISTGRAG 614

Qy   1318 ---EVTGEEETHSPRSV--QYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGV 1372
Db    615 ITLDHTDDKEEQLPRTLAGSSSYLEEE-------------PS------------------ 643

Qy   1373 DDVFHQQSSGETAREVEVDGKGLLPES-VVGEAPTQERGRAADGNTAQTAVNEGDREPVQ 1431
Db    644 DDL------------VCLSDKAIVNRSTILGELKN------------NVVIFEGTL-PKN 678

Qy   1432 SSLVSSPQADIPKVT--QSEVHAQKEVKQEVPLATVSGATPIVD----EKPAPSVTTRGV 1485
Db    679 SVFVSAP--DDPSVTIELSELHAR---------------PVSDFLSMQKPVNIVYTGEV 720

Qy   1486 KIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIA 1545
Db    721 QI-----------------------------------------CQMQNYLD-YLSASLV 737

Qy   1546 TRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHC 1605
Db    738 ACISNLKKYLQDQWLNPGEKFQKIGV------WDNLNNKW----------IVVP------ 775

Qy   1606 LVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYFL 1665
Db    776 --QKKKYAWGLAADVDGNQ--------------------------KTVILNY------- 799

Qy   1666 MPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQ 1725
Db    800 ---------------------------------------------DEH--GMP------- 805

Qy   1726 GATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPM 1785
Db    806 ------------------------------ILEKSYVRL-------------------- 814

Qy   1786 TFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGF 1845
Db    815 ---VVSTDTYLFTVVSMLGYLR---------------------------------------- 833

Qy   1846 HVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFH 1905
Db    834 -----------HLDQ----------------------KKPTATITLVDGV---------- 850

Qy   1906 DAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVG 1965
Db    851 ------------------------------PGCGKTQEILSRF--DANSDLIL---VQ 873

Qy   1966 SSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVR----GMYKRVLVDEVHMMHQGLLQL 2021
Db    874 GREACEMIRRRANDNVPG--SATKENVRTFDSFVMNRKPGKFKTLWVDEGLMVHPGLINF 931

Qy   2022 GVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLS 2081
Db    932 CINISCVSSVYIFGDRKQIPFINRVMNFSIPDNLAKLYYDEIVSRDTTKRCPLDVTHFLN 991

Qy   2082 S------MTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINAD-VYLCMTQLE 2134
Db    992 SVYEKRVMSYSNVQRSLECKMISGKAK------------INDYRSILAEGKLLTFTQED 1038

Qy   2135 KSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTR 2193
Db   1039 KEYL---LKAGFKD--VNTVHEAQGETYRDVNLIRVTATPLTIVSAGSPHVTVALSRHTN 1093

Qy   2194 SLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFAPA 2231
Db   1094 RFVYYTV---VPDVVMTTVQ--KTQCVSNFLLDMYAVA 1126


RESULT   5
V125_AMVLE
ID   V125_AMVLE     STANDARD;      PRT;  1126 AA.
AC   P03589;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-MAR-1992 (Rel. 21, Last annotation update)
DE   125 KDA PROTEIN.
OS   Alfalfa mosaic virus (strain 425 / isolate Leiden).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC   Alfamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83143345; PubMed=6298738;
RA   Cornelissen B.J.C., Brederode F.T., Moormann R.J.M., Bol J.F.;
RT   "Complete nucleotide sequence of alfalfa mosaic virus RNA 1.";
RL   Nucleic Acids Res. 11:1253-1265(1983).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00163; AAA46289.1; -.
DR   PIR; A04197; WMFM12.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   1126 AA;  125827 MW;  BF5A8019B47D4CBF CRC64;


  Query Match             2.9%;  Score 329.5;  DB 1;  Length 1126;
  Best Local Similarity   16.9%;  Pred. No. 3.4e-12;
  Matches  318;  Conservative  191;  Mismatches  488;  Indels  887;  Gaps   63;

Qy    435 TKLCDMFSQ------RDAMIREKPSHRCDVF-----LKPRREREKLRELFPELSIQFSDSV 483
Db     40 TAIGKVFSEAGRAYAQDAL----PSDKGEVLKISFSLDATQQNILRANFPGRRTVFSNSS 95

Qy    484 RSSHPFANAMRSCFNGIFSRRCGNV--CFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVK 541
Db     96 SSSHCFAAAHRLLETDFVYRCFGNTVDSIIDLGGNFVSHMKVKRHNVHCCCPILDARDGA 155

Qy    542 RRINEILFLSTAGGDSYVSS--DLLTEAASKSVSYCSRESQNCDSRADAGFMV-DVYDIS 598
Db    156 RLTERILSLK-----SYVRKHPEIVGEA-----DYCMDTFQKCSRRADYAFAIHSTSDLD 205

Qy    599 PQQVAEAMDKKGALVFDIALMFPVELLYGN-GEVYLEELDTLVKREGDYLAY------NV 651
Db    206 VGELACSLDQKGVMKFICTMMVDADMLIHNEGEIPNFNVRWEIDRKKDLIHFDFIDEPNL 265

Qy    652 GQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGT---E 708
Db    266 G-----YSHRFSLLKHYLTY--------NAVDLGHAAYR-------IERKQDFGGVMVID 305

Qy    709 VTYRSL-----VPSFVGKSLVFIPVVAGSSV-----------SFKTIVLDSDFVDRIYSY 752
Db    306 LTY-SLGFVPKMPHSNGRSCAWYNRVKGQMVVHTVNEGYYHHSYQTAVRRKVLVDKKVLT 364

Qy    753 ALNTIGTFENRTFEYAVGAVRSQKT-------HVITGSRVVHSKVDISPDDMWGLVVAVM 805
Db    365 RVTEVAFRQFRPNADAHSAIQSIATMLSSSTNHTIIGGVTLISGKPLSPDDYIPVATTIY 424

Qy    806 AQAIK------------DRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAK 853
Db    425 YRVKKLYNAIPEMLSLLDKGERLSTDAVLKGSEGPM--------WYSGPTFLSALD---K 473

Qy    854 AMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKN 913
Db    474 VNVPGDF-VAKALLSLPK-----------------------------------RDLKS 495

Qy    914 SRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKA 973
Db    496 ----LFSRS------------ATSHSERTPVRDE----------SPIRC----------- 518
```

```
Qy    974 TVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGI 1033
                                      |  :||                       |: :
Db    519 ----------------------------TDGVFY---------------------PIRM 528

Qy   1034 LETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAA 1093
          |      | :   |  :  |:: |
Db    529 L-------LKCLGSDK--FESVTITD--------------------------------- 545

Qy   1094 NARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVP 1153
                                  |  | :||      : : |   :
Db    546 ----------------------PRSNTETTVD-----LYQSFQKKI------------- 564

Qy   1154 LALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAV 1213
                                       :| | || : |
Db    565 --------------------------ETVFSFILGKIDGP------------------- 578

Qy   1214 PEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLR--YYRDF 1271
                                              |:    || ::|   || ::
Db    579 ---------------------------------------SPLISDPVYFQSLEDVYYAEW 599

Qy   1272 DYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSV 1331
                          : |:    |:|
Db    600 -------------------HQGNAIDASNY---------------------------- 610

Qy   1332 QYTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVD 1391
                                          :  :||:  |:     |:  ||:
Db    611 ----------------------------------ARTLLDDIRKQKEESLKAKAKEVE 634

Qy   1392 GKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVH 1451
               |  :::                                            :||
Db    635 DAQKLNRAIL-------------------------------------------------QVH 647

Qy   1452 AQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQR 1511
          |  |   :                        | ||   | :   :|:  :: :  ||
Db    648 AYLEAHPD----------------------GGKIEGLGLSSQFIAKIPELAIPTPKPL 683

Qy   1512 SLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGE 1571
                 : |: |          ::: :::| :      |  :: :    | :  :|||
Db    684 PEFEKNAETGEIL----------RINPHSDAIL-----EAIDYLKSTSANSIITLNKLG- 727

Qy   1572 GYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNP 1631
                                        |||  |    |  | :  ||:
Db    728 -------------------------------DHC--QWTTKGLDVVWAGDDK-------- 746

Qy   1632 ILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFR 1691

Db    747 ------------------------------------------------------------ 746

Qy   1692 ATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGR 1751
            || |                    :| :        |:       :|:|
Db    747 --RRAF------------------IPKKNTWVGPTARSYPLAKYERAMS---------- 775

Qy   1752 STYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTG-----TSEDAVVEYLKYL 1806
                :  ||     || |  |:      :   :|:|          | :|::  |:
Db    776 -------KDGYVT----LRWDGEVLDANCVRSLSQYEIVFVDQSCVFASAEAIIPSLE-- 822

Qy   1807 AIGIERTYRALLMARNIAVTTAEGVL---KVPN--QVYESLPGFHVYKSGTDLIFHSTQD 1861
                  :|| :  : :||  :||     |  |  |:  |
Db    823 --------KALGLEAHFSVTIVDGVAGCGKTTNIKQIARS------------------- 854

Qy   1862 GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGM 1921
                           |:||| :  |  |    |::    |   |     :  |
Db    855 -------------------SGRDVDLI--LTSNRSSADELKETIDCSPLTKLHYIRTC-- 891

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNL 1981
                                    | :
Db    892 --------------------------DSY------------------------------ 894

Qy   1982 EGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIP 2041
            | ||: | :           :|:: ||  : | ||:       | || : |||  |||
Db    895 --LMSASAVKA----------QRLIFDECFLQHAGLVYAAATLAGCSEVIGFGDTEQIP 941
```

```
Qy   2042 FINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKD 2101
          |::|   |          | |  : |   :| | |  | |             ||
Db    942 FVSRNPSFVFRHHKLTGKVERKLIT---WRSPADATYCL------------EKYFYKNK 985

Qy   2102 KP------VVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVH 2155
          ||      |:||:   || :   |       :||| || ||: :|     || :  : | |
Db    986 KPVKTNSRVLRSIEVVPINSPVSVERNTNALYLCHTQAEKAVLKAQTHLKGCDN-IFTTH 1044

Qy   2156 EAQGKTFSDVVLFRTKKADDSLFTKQPHI------LVGLSRHTRSLVYAALSSKLDDKVG 2209
          ||||||| :|   |  :   || | :  |      || |||| ::  |  ::   || :
Db    1045 EAQGKTFDNVYFCRLTRTSTSLATGRDPINGPCNGLVALSRHKKTFKYFTIAHDSDDVIY 1104

Qy   2210 TYISDASPQSVSDALL-----HTF 2228
               ||   :  |::|     | |
Db    1105 NACRDAG--NTDDSILARSYNHNF 1126
```

RESULT 6
RRPO_TOMS1
ID RRPO_TOMS1 STANDARD; PRT; 1616 AA.
AC Q9YKD6; Q9WJ37;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain S-1) (ToMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
RT virus.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AJ132845; CAB36997.1; -.
DR EMBL; AJ132845; CAB36998.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR001788; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF00978; RNA_dep_RNApol2; 1.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183542 MW; 5DBBFB2FADCC5C0C CRC64;

Query Match 2.7%; Score 313; DB 1; Length 1616;
Best Local Similarity 16.9%; Pred. No. 6.5e-11;
Matches 317; Conservative 204; Mismatches 518; Indels 840; Gaps 63;

```
Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
            |  |    |::: :|     | :  |     || : | |: ||     :|        :
Db      2 AYTQTATSSALLETIRGNNTLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVVSE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
```

```
Db      57 EQTLIATKAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy     522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVSS------DLLTEAASK 570

Db     117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGNKHVPNFQKEAFDRYAEMPNE 176

Qy     571 SVSY-----CSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625

Db     177 VVCHDTFQTCGHSQECYTGRVYAIALHSIYDIPADEFGAALLRKNVHVCYAAFHFSENLL 236

Qy     626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684

Db     237 LEDSHVNLDEINACFQRDGDRLTFSFASESTLNYSHSYSNILKYVCKTYFPASNREVYMK 296

Qy     685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744

Db     297 EF------------------------------------------------------------ 298

Qy     745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802

Db     299 LVTRVNTWFCKFSRIDTF------------------LLYKGVAHKGVD----------- 328

Qy     803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862

Db     329 --------------SEQFYKAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy     863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922

Db     369 MRDMVIVP---------------------------LFDISLET----SKRTRKEVL--- 393

Qy     923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981

Db     394 VLKDFVYTVLNHIRTYQAKAL--TYSNVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy     982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041

Db     446 SL------------LQSLSMTFFLHIKLAVLKDDLL----------------------I 470

Qy    1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097

Db     471 SKFALGPKTVSQHVWDEISL---------------------------------------- 490

Qy    1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157

Db     491 ------------------------------------AFGNAFPSIKERLINRK------ 507

Qy    1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE-- 1215

Db     508 ------------LIKITENALE-------------------------------IRVPDLY 524

Qy    1216 VVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDE 1275

Db     525 VTFHDRLVSEYKMSVDM-------------------PVLDIRKKMEETEEMY------- 557

Qy    1276 GAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTY 1335

Db     558 ---------------------NALSELSVLKNSDKFDVDVFSQMC--------QSL---- 584

Qy    1336 VEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGL 1395

Db     585 ---EVDPMTA------------AKVIVAVMSNESGLTLTFEQPTEANVALAL------- 621

Qy    1396 LPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKE 1455

Db     622 -----------QDSEKASDG---------------ALVVTSRDVEEPSIKGS------M 648

Qy    1456 VKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTI 1515

Db     649 ARGELQLAGLSG------DVPESSYT-----------------RSEEIESLEQFHMAT 683

Qy    1516 NEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY 1575

Db     684 ASSLIHKQMCSI----------VYTGPLKVQQMKN---FIDSLVASLSA---------- 719

Qy    1576 NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTV 1635

Db     720 -----------AVSNLVKILKDTAAID--LETRQKFG--------------------- 743

Qy    1636 NLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SIDEGR 1685

Db     744 -------------------VLDVASKRWLVKPSAKNHAWGVVETHARKYHVALLEHDE-- 782

Qy    1686 ISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG 1745

Db     783 -----------FGI------------------------------------------- 785

Qy    1746 REVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805

Db     786 ----------ITCD-------NWRR-----VAVSP-------------------ESVVY 803

Qy    1806 LAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRV 1865

Db     804 SDMAKLRTLRRLLKDGEPHVSSAKVVL------------------------------- 830

Qy    1866 RDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGES 1925

Db     831 -------------------VDGV----------------------------------- 834

Qy    1926 FKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------VKKRD 1978

Db     835 ------------PGCGK--------------TKEILSRVNFEEDLILVPGRQAAEMIRRR 868

Qy    1979 PNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031

Db     869 ANASGIIVATKDNVRTVDSFLMNYGKGARCQFKRLFIDEGLMLHTGCVNFLVEMSLSDIA 928

Qy    2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTV 2085

Db     929 YVYGDTQQIPYINRVTGFPYPAHFAKLEVDEVETRRTTLRCPADVTHFLNQRYEGHVMCT 988

Qy    2086 RGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143

Db     989 SSEKKSVSQEMVSGAASINPVSKPLKGK--------------ILTFTQ---SD-KEALL 1029

Qy    2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYAA--- 2199

Db    1030 SRG-YTDVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYYTVVM 1088

Qy    2200 --LSSKLDD--KVGTYISD 2214

Db    1089 DPLVSIIRDLERVSSYLLD 1107
```

RESULT 7
RRPO_TOML
ID RRPO_TOML STANDARD; PRT; 1616 AA.
AC P03587; O41352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain L) (ToMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
RA Okada Y.;
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT genome and comparison with the common strain genome.";
RL J. Biochem. 96:1915-1923(1984).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA

```
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02144; CAA26085.1; -.
DR   EMBL; X02144; CAA26082.1; -.
DR   PIR; A04195; WMTM8T.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183564 MW;  A8EC8929B5CF7CAF CRC64;
```

Query Match 2.7%; Score 309; DB 1; Length 1616;
Best Local Similarity 16.8%; Pred. No. 1.1e-10;
Matches 315; Conservative 205; Mismatches 519; Indels 840; Gaps 64;

```
Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
Db      2 AYTQTATSSALLETVRGNNTLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVVSE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
Db     57 EQTLIATKAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVS----------SDLLTE 566
Db    117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGNKHVPNFQKEAFDRYAEMPNE 176

Qy    567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
Db    177 VVCHDTFQTCRHSQECYTGRVYAIALHSIYDIPADEFGAALLRKNVHVCYAAFHFSENLL 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
Db    237 LEDSHVNLDEINACFQRDGDRLTFSFASESTLNYSHSYSNILKYVCKTYFPASNREVYMK 296

Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
Db    297 EF------------------------------------------------------------ 298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
Db    299 LVTRVNTWFCKFSRIDTF------------------LLYKGVAHKGVD----------- 328

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
Db    329 --------------SEQFYKAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
Db    369 MRDMVIVP---------------------------LFDISLET----SKRTRKEVL--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
Db    394 VSKDFVYTVLNHIRTYQAKAL--TYSNVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041

Db    446 SLLQSLSMT--------FF-------------------------LHTKLAVLKDD-LLI 470

Qy   1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097
Db    471 SKFALGPKTVSQHVWDEISL---------------------------------------- 490

Qy   1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157
Db    491 ------------------------------------AFGNAFPSIKERLINRK------ 507

Qy   1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE-- 1215
Db    508 ------------LIKITENALE-------------------------------IRVPDLY 524

Qy   1216 VVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDE 1275
Db    525 VTFHDRLVSEYKMSVDM-------------------PVLDIRKKMEETEEMY------- 557

Qy   1276 GAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTY 1335
Db    558 ---------------------NALSELSVLKNSDKFDVDVFSQMC--------QSL---- 584

Qy   1336 VEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGL 1395
Db    585 ---EVDPMTA-------------AKVIVAVMSNESGLTLTFEQPTEANVALAL------- 621

Qy   1396 LPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKE 1455
Db    622 -----------QDSEKASDG---------------ALVVTSRDVEEPSIKGS------M 648

Qy   1456 VKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTI 1515
Db    649 ARGELQLAGLSG------DVPESSYT------------------RSEEIESLEQFHMAT 683

Qy   1516 NEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY 1575
Db    684 ASSLIHKQMCSI----------VYTGPLKVQQMKN---FIDSLVASLSA----------- 719

Qy   1576 NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTV 1635
Db    720 ------------AVSNLVKILKDTAAID--LETRQKFG--------------------- 743

Qy   1636 NLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SIDEGR 1685
Db    744 -------------------VLDVASKRWLVKPSAKNHAWGVVETHARKYHVALLEHDE-- 782

Qy   1686 ISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG 1745
Db    783 -----------FGI------------------------------------------------ 785

Qy   1746 REVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805
Db    786 ----------ITCD-------NWRR--------------------VAVSSESVV----Y 803

Qy   1806 LAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRV 1865
Db    804 SDMAKLRTLRRLLKDGEPHVSSAKVVL------------------------------- 830

Qy   1866 RDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGES 1925
Db    831 -------------------VDGV------------------------------------ 834

Qy   1926 FKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------VKKRD 1978
Db    835 ------------PGCGK-------------TKEILSRVNFEEDLILVPGRQAAEMIRRR 868

Qy   1979 PNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
Db    869 ANASGIIVATKDNVRTVDSFLMNYGKGARCQFKRLFIDEGLMLHTGCVNFLVEMSLCDIA 928

Qy   2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTV 2085
```

```
Db    929 YVYGDTQQIPYINRVTGFPYPAHFAKLEVDEVETRRTTLRCPADVTHFLNQRYEGHVMCT 988

Qy   2086 RGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
Db    989 SSEKKSVSQEMVSGAASINPVSKPLKGK--------------ILTFTQ---SD-KEALL 1029

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYAA--- 2199
Db   1030 SRG-YADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYYTVVM 1088

Qy   2200 --LSSKLDD--KVGTYISD 2214
Db   1089 DPLVSIIRDLERVSSYLLD 1107


RESULT   8
RRPO_TOMK2
ID   RRPO_TOMK2     STANDARD;      PRT;  1616 AA.
AC   P89676; P90349;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain Kazakh K2) (ToMV) (TMV strain K2).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98012318; PubMed=9454068;
RA   Belenovich E.V., Generozov E.V., Novikov V.K., Zavriev S.K.;
RT   "Properties and structure of the tobacco mosaic virus strain K2
RT   genome.";
RL   Mol. Biol. (Mosk) 31:826-830(1997).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z92909; CAB07439.1; -.
DR   EMBL; Z92909; CAB07438.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183614 MW;  C2F5CE0C8C965336 CRC64;


  Query Match             2.6%;  Score 303;  DB 1;  Length 1616;
  Best Local Similarity   16.7%;  Pred. No. 2.7e-10;
  Matches  314;  Conservative  206;  Mismatches  519;  Indels  840;  Gaps   64;

Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
Db      2 AYTQTATSSALLETVRGNNTLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVVSE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
Db     57 EQTLIATKAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVS----------SDLLTE 566
Db    117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGNKHVPNFQKEAFERYAEMPNE 176

Qy    567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
Db    177 VVCHDTFQTCRHSQECYTGRVYAIALHSIYDIPADEFGAALLRKNVHVCYAAFHFSENLL 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
Db    237 LEDSHVNLDEINACFQRDGDRLTFSFASESTLNHSHSYSNILKYVCKTYFPASNREVYMK 296

Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
Db    297 EF---------------------------------------------------------- 298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
Db    299 LVTRVNTWFCKFSRIDTF------------------LFYKGVAHKGVD----------- 328

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
Db    329 --------------SEQFYKAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
Db    369 MRDMVIVP---------------------------LFDISLET----SKRTRKEVL--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
Db    394 VSKDFVYTVLNHIRTYQAKAL--TYSNVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041
Db    446 SLLQSLSMT--------FF-------------------------LHTKLAVLKDD-LLI 470

Qy   1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097
Db    471 SKFALGPKTVSQHVWDEISL---------------------------------------- 490

Qy   1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157
Db    491 ------------------------------------AFGNAFPSIKERLINRK------ 507

Qy   1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE-- 1215
Db    508 -----------LIKITENALE-----------------------------IRVPDLY 524

Qy   1216 VVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDE 1275
Db    525 VTFHDRLVSEYKMSVDM-------------------PVLDIRKKMEETEEMY------- 557

Qy   1276 GAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTY 1335
Db    558 --------------------NALSELSVLKNSDKFDVDVFSQMC--------QSL---- 584

Qy   1336 VEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGL 1395
Db    585 ---EVDPMTA-------------AKVIVAVMSNESGLTLTFEQPTEANVALAL------- 621

Qy   1396 LPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKE 1455
Db    622 -----------QDSEKASDG---------------ALVVTSRDVEEPSIRGS------M 648

Qy   1456 VKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTI 1515
Db    649 ARGELQLAGLSG------DVPESSYT-----------------RSEEIESLEQFHMAT 683

Qy   1516 NEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY 1575
Db    684 ASSLIHKQMCSI----------VYTGPLKVQQMKN---FIDSLVASLSA---------- 719
```

```
Qy   1576 NGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTV 1635
                    |: :::  :|  :  |  |  : | |
Db    720 ------------AVSNLVKILKDTAAID--LETRQKFG---------------------- 743

Qy   1636 NLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SIDEGR 1685
                          |::|||  : : |       |   ||    :      ||
Db    744 ------------------VLDVASKRWLVKPSAKNHAWGVVETHARKYHVALLEHDE-- 782

Qy   1686 ISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSG 1745
                    ||:
Db    783 -----------FGI---------------------------------------------- 785

Qy   1746 REVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805
                    | |:        | |                          | ::||    |
Db    786 ----------ITCD-------NWRR---------------------VAVSSESVV----Y 803

Qy   1806 LAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRV 1865
            :   || | ||      |::|: ||
Db    804 SDMAKLRTLRRLLKDGEPHVSSAKVVL------------------------------- 830

Qy   1866 RDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGES 1925
                          || |
Db    831 ------------------VDGV------------------------------------- 834

Qy   1926 FKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------VKKRD 1978
                    || ||              |  | : |   ||: :        : :|
Db    835 ------------PGCGK-------------TKEILSRVNFEEDLILVPGRQAAEMIRRR 868

Qy   1979 PNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATGASEG 2031
            |  |:  ||  | | | :|::      |  :||: :||  |:| | :  |  :
Db    869 ANASGIIVATKDNVRTVDSFLMNYGKGARCQFKRLFIDEGLMLHTGCVNFLVEMSLCDIA 928

Qy   2032 LFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTV 2085
            :||  |||:|||   |          : : |     : ||| || : |:       |
Db    929 YVYGDTQQIPYINRVTGFPYPAHFAKLEVDEVETRRTTLRCPADVTHFLNQRYEGHVMCT 988

Qy   2086 RGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
            :|   :::|||     || : |  :                |  ||   || | :|
Db    989 SSEKKSVSQEMVSGAASINPVSKPLKGK--------------ILTFTQ---SD-KEALL 1029

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYAA--- 2199
            :|    | |||| ||:|::|| | |      |:  :  ||:|| |||||:|| |
Db   1030 SRG-YADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYYTVVM 1088

Qy   2200 --LSSKLDD--KVGTYISD 2214
            | | : |  :| :|: |
Db   1089 DPLVSIIRDLERVSSYLLD 1107


RESULT   9
RRPO_TOMK1
ID   RRPO_TOMK1     STANDARD;      PRT;  1616 AA.
AC   Q9Q1T8; Q9Q1T7;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tomato mosaic virus (strain Kazakh K1) (ToMV) (TMV strain K1).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196905; PubMed=10732356;
RA   Belenovich E.V., Novikov V.K., Zavriev S.K.;
RT   "Biological properties and genome structure of the Kazakh isolate K1
RT   of Tobacco Mosaic virus.";
RL   Mol. Biol. (Mosk) 34:172-176(2000).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ243571; CAB62911.1; -.
DR   EMBL; AJ243571; CAB62912.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183606 MW;  4BF9913CE791DE33 CRC64;


  Query Match             2.6%;  Score 302;  DB 1;  Length 1616;
  Best Local Similarity   16.7%;  Pred. No. 3.1e-10;
  Matches  315;  Conservative  204;  Mismatches  516;  Indels  848;  Gaps   65;

Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
            |  |    |::: ::    | :  |     || : | |: ||     :|       :
Db      2 AYTQTATSSALLETVRGNNTLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVVSE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
            :     : :||  | | ::  :  |  |  :||      :  :       :||||:|  |:
Db     57 EQTLIATKAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLS-TAGGDSYVS----------SDLLTE 566
            |    | | | |||:|: |   :     |:||    |: :|           :::  |
Db    117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGNKHVPNFQKEAFDRYAEMPNE 176

Qy    567 AASKSVSYCSRESQNC-DSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
                | || |   |  |  :  :|||   :   |: :|   |   |  |   ||
Db    177 VVCHDTFQTCRHSQECYTGRVYAIALHSIYDIPADEFGAALLRKNVHVCYAAFHFSENLL 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
            :  | |:|::   :|:|| | ::      : | ||:||:  :   :|   |:  |:
Db    237 LEDSHVNLDEINACFQRDGDRLTFSFASESTLNYSHSYSNILKYVCKTYFPASNREVYMK 296

Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
            |:
Db    297 EF---------------------------------------------------------- 298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
            | |: ::    : | ||                :    | |  ||
Db    299 LVTRVNTWFCKFSRIDTF------------------LLYKGVAHKGVD----------- 328

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
                        :  | || | :       : :|:  | |  : :   : |:  |
Db    329 --------------NEQFYKAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
            :  :: :|                              |||| ||     | :|  |
Db    369 MRDMVIVP---------------------------LFDISLET----SKRTRKEVL--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
            : :| :  |:  ::|   | |  |  :| :|:    | |       :: |  |     | |
Db    394 VSKDFVYTVLNHIRTYQAKAL--TYSNVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041
            |:   : |        ||                        ::: : :|:   : :
Db    446 SLLQSLSMT--------FF------------------------LHTKLAVLKDD-LLI 470
```

```
Qy   1042 SKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARR 1097
Db    471 SKFALGPKTVSQHVWDEISL---------------------------------------- 490

Qy   1098 YAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALL 1157
Db    491 ------------------------------------------------------------ 490

Qy   1158 SFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVV 1217
Db    491 ----------------------------------------------AFGNAFPSIKE-- 501

Qy   1218 WRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGA 1277
Db    502 ---RLINRKLIKIT-------ENALEIRVPDL-YV------------------------- 525

Qy   1278 GPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVE 1337
Db    526 ----TFHDRL-----------VSEYKMSVDMPVLDIRKKMEETEE---------MYNALS 561

Qy   1338 EEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLP 1397
Db    562 E-----------------------LSVLKTSDKFDVDVFSQM-----CQSLEVD--PMTA 591

Qy   1398 ESVVGEAPTQERG------RAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVH 1451
Db    592 AKVIVAVMSNESGLTLTFEQPTEANVA-LALQDSEKASDGALVVTSRDVEEPSIRGS--- 647

Qy   1452 AQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQR 1511
Db    648 ---MARGELQLAGLSG------DVPESSYT-----------------RSEEIESLEQF 679

Qy   1512 SLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGE 1571
Db    680 HMATASSLIHKQMCSI----------VYTGPLKVQQMKN---FIDSLVASLSA------- 719

Qy   1572 GYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNP 1631
Db    720 ---------------AVSNLVKILKDTAAID--LETRQKFG------------------ 743

Qy   1632 ILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPC------GFQRTHLHSVN----SI 1681
Db    744 ----------------------VLDVASKRWLVKPSAKNHAWGVVETHARKYHVALLEH 780

Qy   1682 DEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALS 1741
Db    781 DE------------FGI---------------------------------------- 785

Qy   1742 EGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVE 1801
Db    786 -------------ITCD-------NWRR---------------------VAVSSESVV- 802

Qy   1802 YLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQD 1861
Db    803 ---YSDMAKLRTLRRLLKDGEPHVSSAKVVL---------------------------- 830

Qy   1862 GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGM 1921
Db    831 ----------------------VDGV---------------------------------- 834

Qy   1922 VGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMA-------V 1974
Db    835 ---------------PGCGK-------------TKEILSRVNFEEDLILVPGRQAAEM 864

Qy   1975 KKRDPNLEGLNSATTVNSRVV-NFIV------RGMYKRVLVDEVHMMHQGLLQLGVFATG 2027
Db    865 IRRRANASGIIVATKDNVRTVDSFLMNYGKGARCQFKRLFIDEGLMLHTGCVNFLVEMSL 924

Qy   2028 ASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS----- 2082
Db    925 CDIAYVYGDTQQIPYINRVTGFPYPAHFAKLEVDEVETRRTTLRCPADVTHFLNQRYEGH 984

Qy   2083 -MTVRGTEKCYPEKVVSG--KDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMK 2139
Db    985 VMCTSSEKKSVSQEMVSGAASINPVSKPLKGK---------------ILTFTQ---SD-K 1025

Qy   2140 RSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTK-QPHILVGLSRHTRSLVYA 2198
Db    1026 EALLSRG-YADVHTVHEVQGETYADVSLVRLTPTPVSIIARDSPHVLVSLSRHTKSLKYY 1084

Qy   2199 A-----LSSKLDD--KVGTYISD 2214
Db    1085 TVVMDPLVSIIRDLERVSSYLLD 1107
```

RESULT 10
RRPO_CGMVS
ID RRPO_CGMVS STANDARD; PRT; 1648 AA.
AC P19523; Q83208; P89877; P90356;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
OS Cucumber green mottle mosaic virus (watermelon strain SH) (CGMMV), and
OS Cucumber green mottle mosaic virus (watermelon strain W) (CGMMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SH;
RX MEDLINE=91311400; PubMed=1856687;
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA Sato T., Motoyoshi F., Nishiguchi M.;
RT "The complete nucleotide sequence of cucumber green mottle mosaic
RT virus (SH strain) genomic RNA.";
RL J. Gen. Virol. 72:1487-1495(1991).
RN [2]
RP REVISIONS.
RC STRAIN=SH;
RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
RA Sato T., Motoyoshi F., Nishiguchi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 1496-1648 FROM N.A.
RC STRAIN=W;
RX MEDLINE=89073773; PubMed=3201760;
RA Saito T., Imai Y., Meshi T., Okada Y.;
RT "Interviral homologies of the 30K proteins of tobamoviruses.";
RL Virology 167:653-656(1988).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN
CC CODONS FOR LYS-1144 AND GLN-1146.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; D12505; BAA18895.1; -.
DR EMBL; D12505; BAA18896.1; -.
DR EMBL; J04322; AAA46382.1; -.
DR PIR; JQ1157; WMTMS2.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR001788; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF00978; RNA_dep_RNApol2; 1.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1648 RNA-DIRECTED RNA POLYMERASE.

```
FT   CHAIN        1   1144       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND    863    870       ATP (POTENTIAL).
SQ   SEQUENCE   1648 AA;  186549 MW;  1D1AFFEEE7B65595 CRC64;


  Query Match             2.6%;  Score 298;  DB 1;  Length 1648;
  Best Local Similarity   17.4%;  Pred. No. 5.7e-10;
  Matches  334;  Conservative  192;  Mismatches  509;  Indels  884;  Gaps   72;

Qy     405 NQEEVYSS----VVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFL 460
           || :  :|    :|| : :: |  :    | | |                :|       :
Db      10 NQRDAAASGRNNLVSQLASKRVYDEAVRSLDHQD--------------RRPKMNFSRVV 54

Qy     461 KPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRC--GNVCFFDIGGSFT 518
                 : : :|| || |: :  | |  |  :|         :   |: | :||||::|
Db      55 STEHTRLVTDAYPEFSISFTATKNSVHSLAGGLRLLELEYMMMQVPYGSPC-YDIGGNYT 113

Qy     519 YHVKAGHVNCHVCNPVLDVKDVKRRI--NEILF-------------LSTAGGDSYVSSD 562
           |:  |    | ||| ||:||| | :  |:::               | |   |::   |
Db     114 QHLFKGRSYVHCCNPCLDLKDVARNVMYNDMITQHVQRHKGSGGCRPLPTFQIDAFRRYD 173

Qy     563 LLTEAASKSVSYCSRESQNCDSRADAG------FMVDVYDISPQQVAEAMDKKGALVFDI 616
               :|     ||   | |      :|       :  :|||    :  |: :|   |
Db     174 -----SSPCAVTCSDVFQECSYDFGSGRDNHAVSLHSIYDIPYSSIGPALHRKNVRVCYA 228

Qy     617 ALMFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVR 675
           |  |   || |:    |  :    : :|| : :   :   : | ||  |:      :|
Db     229 AFHFSEALLLGSPVGNLNSIGAQFRVDGDDVHFLFSEESTLHYTHSLENIKLIVMRTYFP 288

Qy     676 TSSGNVFKLEYEGYRCGYHHLTMCRA-----QKSPG------TEVTYRSLVPSFVGKSL- 723
               |:  |:   |       : ||      || |      :|    :  | |  |:
Db     289 ADDRFVYIKEFMVKRVDTFFFRLVRADTHMLHKSVGHYSKSKSEYFALNTPPIFQDKATF 348

Qy     724 ----------VFIPV-----VAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFENR--TFE 766
                    | ||          :|    :::|:|||  |    :| | |::|:   ::
Db     349 SVWFPEAKRKVLIPKFELSRFLSGNVKISRMLVDADFVHTI----INHISTYDNKALVWK 404

Qy     767 YAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIR--SYNFIKA 824
           | | :: ||     | |: ::  | :  :  ::    :| |   |   |   :
Db     405 NVQSFVESIRSRVIVNGVSVKSEWNVPVDQLTDISFSIFL-LVKVRKVQIELMSDKVVIE 463

Qy     825 SEGSL---AGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSV 881
           : | |   |   |   : :|||      ||
Db     464 ARGLLRRFADSLKSAVEGLGDC------VY----------------------------- 487

Qy     882 VVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKP 941
                 | |   | || |                       | :||:          |
Db     488 ----------DALVQTGWFDTS--------------------SDELKVL--------LP 508

Qy     942 LPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITSFFS 1001
           |    | |                |:          |||: | :     |    | : :|
Db     509 EPFMTFSDYL-------------EGMYEADAKIERESVSELLASGDDLFKKIDEIRNNYS 555

Qy    1002 TGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLNDFT 1061
                |:  |                                |   :||  |:::
Db     556 --------GVEFD-----------------------------VEKFQEFCKELNV---- 573

Qy    1062 TFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLRGLT 1121
                | :|||                                : :: || ||
Db     574 -----NPMLIG-----------------------------HVIEAIFSQKAG------ 591

Qy    1122 SSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRH 1181
                         ||||
Db     592 ------------------VTVT----------------------------------- 595

Qy    1182 LMFFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGL 1241
           |||||                  ||:   | ||:
Db     596 -----GLGTL--------------------SPEM---GASVA---------------- 609

Qy    1242 LNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSS 1301
           |: |                                                  |  :
Db     610 LSST-------------------------------------------------SVDTCED 620

Qy    1302 YDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGSGTADA 1361
              ||  :: |:       |   :::||       |:  |:|  |
Db     621 MDVTEDMEDI-------VLMADKSHS-------YMSPEMARWA---------------- 649

Qy    1362 MAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPT-----QERGRAADGN 1416
                 || |                    : || | |  || : |      |:|:|
Db     650 -----DVKYG-------------------NNKGALVEYKVGTSMTLPATWAEKGKAV--- 682

Qy    1417 TAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKP 1476
                       |:    |  ||           | : :|           |:
Db     683 -----------LPLSGICVRKPQ-----------FSKPLDEE-----------DDLR 706

Qy    1477 APSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQL 1536
              ::    |  :   | :  |      ::  | |
Db     707 LSNMNFFKVSDLKLKKTITPVVYTGTIRERQMK------------------------- 739

Qy    1537 DVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAI 1596
                          ::| |   ||   | ||     |   | | |    | : |
Db     740 ---------------NYIDYL---SASLGSTLG-----NLERIVRSDWNGTEESMQTFG 775

Qy    1597 KYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVI 1656
           |                      | |:|
Db     776 LY---------------------DCEKC------------------------------ 782

Qy    1657 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPG 1716
                 : |:|   :: |  :|       :  :   || :|           :| ::| |
Db     783 -----KWLLLPA--EKKHAWAV-------VLASDDTTRIIF-----------LSYDES-G 816

Qy    1717 VPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVM 1776
           |           |  |   |  :  |  :|     ||:      :|  | |
Db     817 SP------------IIDKKNWKRFAVCSETKV-----YSV------IRSLEVL------- 846

Qy    1777 ALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPN 1836
                           :::|:|:
Db     847 ------------------NKEAIVD---------------------------------- 853

Qy    1837 QVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLF 1896
                 || |:          :  ||:
Db     854 ------PGVHI----------TLVDGV------------------------------- 864

Qy    1897 VCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLV------DEF 1950
                                                    || |||  ::       :
Db     865 ------------------------------------------PGCGKTAEIIARVNWKTDL 883

Qy    1951 VKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIV--RGMYK--RV 2006
           | :|   |              :::|   |     ||  | |  :  |  | ::|  |
Db     884 VLTPGREAA------------AMIRRRACALHKSPVATNDNVRTFDSFVMNRKIFKFDAV 931

Qy    2007 LVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKK------ 2060
              |||  |:| |||    :  :|  :   |||  |||||||  |   |     ||:
Db     932 YVDEGLMVHTGLLNFALKISGCKKAFVFGDAKQIPFINR--VMNFD----YPKELRTLIV 985

Qy    2061 ESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSK---------R 2111
           ::|     ::||| ||   |:::           |       ||| |:           |
Db     986 DNVERRYVTHRCPRDVTSFLNTI----------YKAAVATTSPVVHSVKAIKVSGAGILR 1035

Qy    2112 PIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTK 2171
           |     :: :|   : :  || :|  : :|         | |||| ||:|| :  : |
Db    1036 P-----ELTKIKGKI-ITFTQSDKQSLIKS----*GYNDVNTVHEIQGETFEETAVVRAT 1084

Qy    2172 KADDSLFTK-QPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTFA 2229
                 |  :  ||:|| |:|||:::||  :   : | | : |:|   : |  ::|  ||
Db    1085 PTPIGLIARDSPHVLVALTRHTKAMVYYTV---VFDAVTSIIADV--EKVDQSILTMFA 1138


RESULT  11
RRPO_PMMVJ
ID   RRPO_PMMVJ     STANDARD;      PRT;  1612 AA.
AC   P89657; P90347;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
```

```
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Pepper mild mottle virus (strain Japan) (PMMV-J).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kirita M., Akutsu K., Watanabe Y., Tsuda S.;
RT   "Nucleotide sequence of the Japanese isolate of pepper mild mottle
RT   tobamovirus.";
RL   Ann. Phytopathol. Soc. Jpn. 0:0-0(1997).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB000709; BAA19167.1; -.
DR   EMBL; AB000709; BAA19166.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltrans; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1612       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1117       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     835    842       ATP (POTENTIAL).
SQ   SEQUENCE   1612 AA;  183303 MW;  E7E66CE8ED81FE4C CRC64;


  Query Match             2.6%;  Score 297;  DB 1;  Length 1612;
  Best Local Similarity   22.8%;  Pred. No. 6.3e-10;
  Matches  195;  Conservative  114;  Mismatches  301;  Indels  244;  Gaps   40;

Qy   1508 PKQRSLTI---------NEGK--AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVD 1556
          || : : |         ||||  | |:: : :     |   :    : |  ::|  :||:
Db    366 PKMKDMVIVPLFDVSLQNEGKRLARKEVMVSKDFVYTVLNHIRTYQSKALTYANVLSFVE 425

Qy   1557 NLKGR----------------------SAVFF--SKLG--------EGYTYNGGSHVSSG 1584
          ::: |                      |  ||  :||         : :  :  |
Db    426 SIRSRVIINGVTARSEWDVDKALLQSLSMTFFLQTKLAMLKDDLVVQKFQVHSKSLTEYV 485

Qy   1585 W-----------PRALEDILT----------AIKYPSV---FDHCLVQKYKMGGGVP-- 1617
          |           |   | ::             || | :   |   ||::||    :|
Db    486 WDEITAAFHNCFPTIKERLINKKLITVSEKALEIKVPDLYVTFHDRLVKEYKSSVEMPVL 545

Qy   1618 -FHADDEECYPSDNPILTVNLVGKAN------FSTKCRKGG-------KVMVINVASGDY 1663
           ||     | :  ::::  ::       ||  |   |        ||||  |:
Db    546 DVKKSLEEAEVMYNALSEISILKDSDKFDVDVFSRMCNTLGVDPLVAAKVMVAVVS---- 601

Qy   1664 FLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQ 1723
                          :|  ::|||         |   ||              |    :
Db    602 ------------------NESGLTLTFERPTEA-NVALALQ------------PTITSK 629

Qy   1724 SQGATRTITPKSGGKALSEGSGREVKGRSTYSI------WCEQDYVRKCEWLRADNPVMA 1777
           :|: : ::   |     | | :||  :|  |:          :: |  |       : :
Db    630 EEGSLKIVSSDVG-----ESSIKEVVRKSEISMLGLTGNTVSDEFQRSTE-------IES 677

Qy   1778 LEPDYTPMTFEVVKTGT---SEDAVVEYLKYLAIGIERTY-RALLMARNIAVTTAEGVLK 1833
          |:       | :| | |    :   : |   | :   : |  :|: : : ||:  : ::|
Db    678 LQ------QFHMVSTETIIRKQMHAMVYTGPLKVQQCKNYLDSLVASLSAAVSNLKKIIK 731

Qy   1834 VPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGI---FTKGKDVDAVVA 1890
                          |  |   |::   |    | :  :| :    :||     |:

Db    732 ----------------DTAAIDLETKEKFGV----YDVCLKKWLVKPLSKGHAWGVVM- 769

Qy   1891 LGDNLFVCDDILVFHDAINLIGALKVARCG----MVGESFKSFEYKCY------------ 1934
            |: : |   |: :|  |::       ||     |  | :|  |
Db    770 --DSDYKCFVALLTYDGENIV-------CGETWRRVAVSSESLVYSDMGKIRAIRSVLKD 820

Qy   1935 NAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKK-------RDPNLEGLNSA 1987
            |       ||||  |     |  | : |   ||: :   |       |  |  ||  |
Db    821 GEPHISSAKVTLVDG-VPGCGKTKEILSRVNFDEDLVLVPGKQAAEMIRRRANSSGLIVA 879

Qy   1988 TTVNSRVV-NFIV---RG--MYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIP 2041
          |  | | |  :|::   ||   |||: :||  |:| | :   |  :  ||   :||  |||
Db    880 TKENVRTVDSFLMNYGRGPCQYKRLFLDEGLMLHPGCVNFLVGMSLCSEAFVYGDTQQIP 939

Qy   2042 FINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTVRGTEKCYPEK 2095
          :|||   |     :   : ::|     : ||| |: : |:       |      :    :
Db    940 YINRVATFPYPKHLSQLEVDAVETRRTTLRCPADITFFLNQKYEGQVMCTSSVTRSVSHE 999

Qy   2096 VVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVH 2155
          |: |    |:  :||   |             :  || :||     |  :| |  | |||
Db   1000 VIQG--AAVMNPVSKPLKGKV----------ITFTQSDKS----LLLSRGYE-DVHTVH 1041

Qy   2156 EAQGKTFSDVVLFRTKKADDSLFTKQ-PHILVGLSRHTRSLVY-----AALSSKLDDK-- 2207
          | ||:|| || | |       : :|| ||:|| |||||||: |      |: | | |
Db   1042 EVQGETFEDVSLVRLTPTPVGIISKQSPHLLVSLSRHTRSIKYYTVVLDAVVSVLRDLEC 1101

Qy   2208 VGTYISDASPQSVS 2221
          | :|: |     ||
Db   1102 VSSYLLDMYKVDVS 1115


RESULT  12
RRPO_PMMVS
ID   RRPO_PMMVS     STANDARD;      PRT;  1612 AA.
AC   P29098; Q84924;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Pepper mild mottle virus (strain Spain) (PMMV-S).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92113528; PubMed=1765765;
RA   Alonso E., Garcia-Luque I., de la Cruz A., Wicke B.,
RA   Avila-Rincon M.J., Serra M.T., Castresana C., Diaz-Ruiz J.R.;
RT   "Nucleotide sequence of the genomic RNA of pepper mild mottle virus,
RT   a resistance-breaking tobamovirus in pepper.";
RL   J. Gen. Virol. 72:2875-2884(1991).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81413; AAB02334.1; -.
DR   EMBL; M81413; AAB02335.1; -.
DR   PIR; JQ1312; WMTMPV.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
```

```
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1612       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1117       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     835    842       ATP (POTENTIAL).
SQ   SEQUENCE   1612 AA;  183265 MW;  465677F4924802BF CRC64;


  Query Match             2.6%;  Score 297;  DB 1;  Length 1612;
  Best Local Similarity   22.8%;  Pred. No. 6.3e-10;
  Matches  195;  Conservative  114;  Mismatches  301;  Indels  244;  Gaps   40;

Qy   1508 PKQRSLTI---------NEGK--AGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVD 1556
          || : : |         |||| | |:: : :     |   :    : | ::|  :||:
Db    366 PKMKDMVIVPLFDVSLQNEGKRLARKEVMVSKDFVYTVLNHIRTYQSKALTYANVLSFVE 425

Qy   1557 NLKGR---------------------SAVFF--SKLG--------EGYTYNGGSHVSSG 1584
          ::: |                       |  ||  :||         : :  :  |
Db    426 SIRSRVIINGVTARSEWDVDKALLQSLSMTFFLQTKLAMLKDDLVVQKFQVHSKSLTEYV 485

Qy   1585 W----------PRALEDILT----------AIKYPSV---FDHCLVQKYKMGGGVP-- 1617
          |          |   | ::           || | :   |   ||::||    :|
Db    486 WDEITAAFHNCFPTIKERLINKKLITVSEKALEIKVPDLYVTFHDRLVKEYKSSVEMPVL 545

Qy   1618 -FHADDEECYPSDNPILTVNLVGKAN------FSTKCRKGG-------KVMVINVASGDY 1663
           ||     | :  :::::  ::      ||  |   |       ||||  |:
Db    546 DVKKSLEEAEVMYNALSEISILKDSDKFDVDVFSRMCNTLGVDPLVAAKVMVAVVS---- 601

Qy   1664 FLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQ 1723
                        :|  ::|||        |   ||              |    :
Db    602 ------------------NESGLTLTFERPTEA-NVALALQ------------PTITSK 629

Qy   1724 SQGATRTITPKSGGKALSEGSGREVKGRSTYSI------WCEQDYVRKCEWLRADNPVMA 1777
           :|: : ::   |     | | :||  :|  |:         :: |  |       : :
Db    630 EEGSLKIVSSDVG-----ESSIKEVVRKSEISMLGLTGNTVSDEFQRSTE-------IES 677

Qy   1778 LEPDYTPMTFEVVKTGT---SEDAVVEYLKYLAIGIERTY-RALLMARNIAVTTAEGVLK 1833
          |:       | :| | |    :   : |   | :   : |  :|: : : ||:  : ::|
Db    678 LQ------QFHMVSTETIIRKQMHAMVYTGPLKVQQCKNYLDSLVASLSAAVSNLKKIIK 731

Qy   1834 VPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGI---FTKGKDVDAVVA 1890
                        |  |   |::   |    | :  :| :    :||     |:
Db    732 -----------------DTAAIDLETKEKFGV----YDVCLKKWLVKPLSKGHAWGVVM- 769

Qy   1891 LGDNLFVCDDILVFHDAINLIGALKVARCG----MVGESFKSFEYKCY------------ 1934
            |: : |   |: :|  |::       ||     |  | :|  |
Db    770 --DSDYKCFVALLTYDGENIV-------CGETWRRVAVSSESLVYSDMGKIRAIRSVLKD 820

Qy   1935 NAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKK-------RDPNLEGLNSA 1987
            |       ||||  |     |  | : |   ||: :   |        |  |  ||  |
Db    821 GEPHISSAKVTLVDG-VPGCGKTKEILSRVNFDEDLVLVPGKQAAEMIRRRANSSGLIVA 879

Qy   1988 TTVNSRVV-NFIV---RG--MYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIP 2041
          |  | | | :|::   ||   |||: :||  |:| | :   |  :  ||   :||  |||
Db    880 TKENVRTVDSFLMNYGRGPCQYKRLFLDEGLMLHPGCVNFLVGMSLCSEAFVYGDTQQIP 939

Qy   2042 FINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------MTVRGTEKCYPEK 2095
          :|||   |     :   : ::|     : ||| |: : |:        |      :    :
Db    940 YINRVATFPYPKHLSQLEVDAVETRRTTLRCPADITFFLNQKYEGQVMCTSSVTRSVSHE 999

Qy   2096 VVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVH 2155
          |: |    |:  :||    |            :  || :||     |  :| |  | |||
Db    1000 VIQG--AAVMNPVSKPLKGKV----------ITFTQSDKS----LLLSRGYE-DVHTVH 1041

Qy   2156 EAQGKTFSDVVLFRTKKADDSLFTKQ-PHILVGLSRHTRSLVY-----AALSSKLDDK-- 2207
          | ||:|| || | |       : :|| ||:|| |||||||: |      |: | | |
Db    1042 EVQGETFEDVSLVRLTPTPVGIISKQSPHLLVSLSRHTRSIKYYTVVLDAVVSVLRDLEC 1101

Qy   2208 VGTYISDASPQSVS 2221
          | :|: |     ||
Db    1102 VSSYLLDMYKVDVS 1115
```

```
RESULT  13
RRPO_TMVRA
ID   RRPO_TMVRA     STANDARD;      PRT;  1616 AA.
AC   Q98745; O41340;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tobacco mosaic virus (strain Rakkyo) (TMV-R).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96265021; PubMed=8678834;
RA   Chen J., Watanabe Y., Sako N., Ohsima K., Okada Y.;
RT   "Complete nucleotide sequence and synthesis of infectious in vitro
RT   transcripts from a full-length cDNA clone of a rakkyo strain of
RT   tobacco mosaic virus.";
RL   Arch. Virol. 141:885-900(1996).
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D63809; BAA09876.1; -.
DR   EMBL; D63809; BAA09877.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN         1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN         1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND     833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183579 MW;  AC7CF5429EBAD49A CRC64;


  Query Match             2.5%;  Score 288;  DB 1;  Length 1616;
  Best Local Similarity   16.8%;  Pred. No. 2.3e-09;
  Matches  316;  Conservative  198;  Mismatches  524;  Indels  842;  Gaps   62;

Qy    403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
          | |    |::: ::     | :  |      || : | |: ||     :|       :
Db      2 AYTQTATTSALLDTVRGNNSLVNDLAKRRLYDTAV-DEFNARD----RRPKVNFSKVISE 56

Qy    463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
          :       :||  | | ::  : |  |  :||     :  :       :||||:|  |:
Db     57 EQTLIATRAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy    522 KAGHVNCHVCNPVLDVKDVKRRINE----ILFLSTAGGDSYVSSDLLTEAASKSVSY--- 574
            |    | | | |||:|: |    :      |:||          :  ||  :   :
Db    117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLDRGGKTVPNFQKEAFDRYAEFPED 176

Qy    575 --CSRESQNCD-------SRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELL 625
            |    | |:         :  |  :  :|||   :   |: :|       |  |   ||
Db    177 AVCHNTFQTCEHQPMQQSGKVYAIALHSIYDIPADEFGAALLRKNVHTCYAAFHFSENLL 236

Qy    626 YGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVFKL 684
          :  | |:|::    |:|| | ::      : | |||||:  :   :|   |:  |:
Db    237 LEDSYVNLDEINACFSRDGDKLTFSFASESTLNYCHSFSNILKYVCKTYFPASNREVYMK 296
```

```
Qy    685 EYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSD 744
          |:
Db    297 EF------------------------------------------------------------298

Qy    745 FVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVV 802
          | |: ::    : | ||             :    | |
Db    299 LVTRVNTWFCKFSRIDTF-----------------LLYKGVAH--------------- 324

Qy    803 AVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNV 862
                  ||: |  |  | | :       : :|:  | |  : :   : |:  |
Db    325 ----------KSVDSEQFYTAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWFPK 368

Qy    863 LETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRA 922
          :  :: :|                              |||| ||     | ||  |
Db    369 MRDMVIIP--------------------------LFDISLET----SKRSRKEVL--- 393

Qy    923 IVEDSI-KVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSK 981
          : :| :  |:  ::|   | |  |  :| :|:    | |      :: |  |     | |
Db    394 VSKDFVFTVLNHIRTYQAKAL--TYANVLSFVESIRSRV------IINGVTARSEWDVDK 445

Qy    982 GLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAV 1041
          |:   : |        ||                          ::: : :|:   : :
Db    446 SLLQSLSMT--------FF-------------------------LHTKLAVLKDD-LLI 470

Qy   1042 SKVVAGTK----EFWSEVSL---NDFTTFV--LRNKVLIGIFVASLGAAPIAWKYRRGIA 1092
          ||   |:|      | |:||   | | :    | || ||
Db    471 SKFSLGSKTVCQHVWDEISLAFGNAFPSVKERLLNKKLI---------------------- 509

Qy   1093 ANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQV 1152
          | || : |                                                :|
Db    510 ----RAAGDALEI-------------------------------------------KV 520

Qy   1153 PLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCA 1212
          |   ::|
Db    521 PDLYITF-------------------------------------------------- 527

Qy   1213 VPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFD 1272
                                                          :|  :  |:
Db    528 ---------------------------------------------HDRLVAEYK--- 536

Qy   1273 YDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQ 1332
                               |:  : : |:   :            |||     |
Db    537 ----------------------SSVDMPALDIRKRM-------------EET----EVM 556

Qy   1333 YTYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDG 1392
          |  : |                          :: :    |   ||| |       : :|||
Db    557 YNALSE----------------------LSVLRESDKFDVDVFSQM-----CKSLEVD- 587

Qy   1393 KGLLPESVVGEAPTQERG------RAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVT 1446
            :    |:    : | |      |  : | || |: : ::    : :|:| : : | :
Db    588 -PMTAAKVIVAVMSNESGLTLTFERPTEANVAQ-ALQDQEKASEGALVVTSREVEEPSMK 645

Qy   1447 QSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVE 1506
            |       : |: ||  :|  |                        :  :  ::|
Db    646 GS------MARGELQLAGFAGDHP------------------------ESSYSRNEEIE 674

Qy   1507 QPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFF 1566
            :|  :   :    ||:              ||       :  |   |:|:|    :
Db    675 SLEQFHMATADSLIRKQMSSI----------VYTGPIKVQQMKN---FIDSLVASLSA-- 719

Qy   1567 SKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECY 1626
                            |: :::  :|  :  |  |  : | |
Db    720 --------------------AVSNLVKILKDTAAID--LETRQKFG------------ 743

Qy   1627 PSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRI 1686
                                  |::|||  : : |
Db    744 --------------------------VLDVASRKWLIKP------------------- 756

Qy   1687 SLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGR 1746
                                                 | ||    : |   |
Db    757 ---------------------------------------------TAKSHAWGVVETHAR 771

Qy   1747 EVKGRSTYSI----WCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEY 1802
          :      | : .  : ||  |   :| |                          | ::||
Db    772 K------YHVALLEYDEQGIVTCDDWRR--------------------VAVSSESVV-- 802

Qy   1803 LKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDG 1862
            |  :   || | ||      |: |: ||
Db    803 --YSDMAKLRTLRRLLRDGEPHVSNAKVVL---------------------------- 830

Qy   1863 LRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMV 1922
                                 || |
Db    831 ----------------------VDGV---------------------------------- 834

Qy   1923 GESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKK------ 1976
                       || ||              |  | : |   ||: :   |
Db    835 --------------PGCGK-------------TKEILSRVNFDEDLILVPGKQAAEMI 865

Qy   1977 -RDPNLEGLNSAT-----TVNSRVVNF--IVRGMYKRVLVDEVHMMHQGLLQLGVFATGA 2028
           |  |  |:  ||     ||:| ::||    |  :||: :||  |:| | :    |  :
Db    866 RRRANSSGIIVATKDNVRTVDSFMMNFGKTTRCQFKRLFIDEGLMLHTGCVNFLVAMSLC 925

Qy   2029 SEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------ 2082
                :||  |||:|||   |         : : |     : ||| || : |:
Db    926 DVAYVYGDTQQIPYINRVSGFPYPAHFSKLEVDEVETRRTTLRCPADVTHYLNRRYEGFV 985

Qy   2083 MTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSL 2142
          ::    :|   :::|||    |:  :||   |              |  ||   || | :|
Db    986 VSTSSVKKSVSQEMVSG--AAVINPISKPLHGK----------ILTFTQ---SD-KEAL 1028

Qy   2143 KGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFT-KQPHILVGLSRHTRSLVYAA-- 2199
            :|  : | |||| ||:|:||| | |       |:      ||:|| ||||| || |
Db    1029 LSRG-YSEVHTVHEVQGETYSDVSLVRLTPTPISIIAGDSPHVLVALSRHTCSLKYYTVV 1087

Qy   2200 ---LSSKLDD--KVGTYISD 2214
              | | : |  |: :|: |
Db    1088 MDPLVSIIRDLEKLSSYLLD 1107
```

RESULT 14
V1A_BBMV
ID V1A_BBMV STANDARD; PRT; 966 AA.
AC Q00020;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 1A PROTEIN [INCLUDES: HELICASE; METHYLTRANSFERASE].
OS Broad bean mottle virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
OC Bromovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAWDEN;
RX MEDLINE=92074218; PubMed=1962437;
RA Dzianott A.M., Bujarski J.J.;
RT "The nucleotide sequence and genome organization of the RNA-1 segment
RT in two bromoviruses: broad bean mottle virus and cowpea chlorotic
RT mottle virus.";
RL Virology 185:553-562(1991).
CC -!- FUNCTION: MAY BE INVOLVED IN THE REPLICATION OF THE VIRUS.
CC CONTAINS AN HELICASE DOMAIN AND A METHYLTRANSFERASE DOMAIN. THE
CC METHYLTRANSFERASE DOMAIN IS PROBABLY INVOLVED IN VIRAL RNA
CC CAPPING.
CC -!- SIMILARITY: TO 1A PROTEIN FROM CCMV, CMV, PSV AND TAV.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M65138; AAA42740.1; -.
DR PIR; A41699; P1BVBB.
DR INTERPRO; IPR000606; -.

```
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Helicase; ATP-binding; Transferase; Methyltransferase.
FT   NP_BIND     690    697       ATP (POTENTIAL).
SQ   SEQUENCE   966 AA;  109621 MW;  DF592681D7231C8D CRC64;


  Query Match             2.5%;  Score 286.5;  DB 1;  Length 966;
  Best Local Similarity   30.9%;  Pred. No. 1.2e-09;
  Matches  105;  Conservative   46;  Mismatches  124;  Indels   65;  Gaps   13;

Qy    1920 GMVGESFKSFEY------KCYNAP----------------------PGGGKTTTLVDEFV 1951
           | :|| |::  |       | ||                        | |||| : : |
Db     646 GGLGEKFENELYVVDKTCSCANAKAIADACKKVSAPTCSVVMVDGVAGCGKTTAIKETF- 704

Qy    1952 KSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVV----NFIVRGMYK--R 2005
                 :||| |:||:  |:        |  | : |  :||    : |: |: :  |
Db     705 -RFEKDIIVTANRKSAEDVRKAI-------FGDASDSEVALKVVRTADSAIMHGLPECHR 756

Qy    2006 VLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMD-CAVFVPKKESVV 2064
           :|||  ::| | |        || | |||  || | :|:  |||  | :   |:: |
Db     757 LLVDEAGLLHYGQLLAVADLCKCSEVLAFGDTEQISFKSRDATFRMKYCNIEYDKRDIV- 815

Qy    2065 YTSKSYRCPLDVCYLLSSMTVRGTEKC------YPEKVVSGKDKPVVRSLSKRPIGTTDD 2118
             ||::||| ||   :|:: :    ||       |  | | |   | :|||  | : :
Db     816 --SKTFRCPQDV---VSAVKIL-KRKCANRSSKYNGWVSSSK---VEKSLSKSRIVSINQ 866

Qy    2119 VAEINADVYLCMTQLEKSDMKRSLKGKG-----KETPVMTVHEAQGKTFSDVVLFRTKKA 2173
           |:      || ||: :|: :    |  |      |: : ||||||||    ||| | |
Db     867 VSMEKHKFYLTMTEADKAALCSRAKDVGLDKTWVESNMETVHEAQGKAVDHVVLVRLKST 926

Qy    2174 DDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYIS 2213
              ||  : : || |:|| |:  |       | :  |::
Db     927 KCDLFKSEEYCLVALTRHKRTFEYLYNGDLGGDLISFYVT 966


RESULT  15
RRPO_TMVB
ID   RRPO_TMVB      STANDARD;      PRT;  1616 AA.
AC   Q9YQ08; Q9WKD4;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE   METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS   Tobacco mosaic virus (strain B935A) (TMV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xue C., Zhou X., Chen Q., Qi Y., Li D.;
RT   "Complete nucleotide sequence and genome organization of tobacco
RT   mosaic virus isolated from Vicia faba.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC       RNA REPLICATION.
CC   -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC       CAPPING AND AN RNA HELICASE.
CC   -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC       BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ011933; CAA09875.1; -.
DR   EMBL; AJ011933; CAA09874.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   CHAIN       1   1616       RNA-DIRECTED RNA POLYMERASE.
FT   CHAIN       1   1116       METHYLTRANSFERASE/RNA HELICASE.
FT   NP_BIND   833    840       ATP (POTENTIAL).
SQ   SEQUENCE   1616 AA;  183296 MW;  99A63B441290E6CF CRC64;


  Query Match             2.5%;  Score 285;  DB 1;  Length 1616;
  Best Local Similarity   16.9%;  Pred. No. 3.5e-09;
  Matches  318;  Conservative  196;  Mismatches  525;  Indels  840;  Gaps   62;

Qy     403 ALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKP 462
           |  |    |::: ::     | :  |     || : : |: ||     :|       :
Db       2 AYTQTATTSALLDTVRGNNSLVNDLAKRRLYDTAV-EEFNARD----RRPKVNFSKVISE 56

Qy     463 REREKLRELFPELSIQFSDSVRSSHPFANAMRSC-FNGIFSRRCGNVCFFDIGGSFTYHV 521
             :       :||  | | ::  : |  |  :||     :  :       :||||:|  |:
Db      57 EQTLIATRAYPEFQITFYNTQNAVHSLAGGLRSLELEYLMMQIPYGSLTYDIGGNFASHL 116

Qy     522 KAGHVNCHVCNPVLDVKDVKRRINE---ILFLS--TAGGDSY-----VSSDLLTEAASK 570
             |    | | | |||:|: |   :     |:||    || :         : |   |
Db     117 FKGRAYVHCCMPNLDVRDIMRHEGQKDSIELYLSRLERGGKTVPNFQKEAFDRYAEIPED 176

Qy     571 SVSYCSRESQNCD-------SRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVE 623
           :|  |    | |:         |  |  :  :|||   :    |: :|       |  |
Db     177 AV--CHNTFQTCEHQPMQQSGRVYAIALHSIYDIPADEFGAALLRKNVHTCYAAFHFSEN 234

Qy     624 LLYGNGEVYLEELDTLVKREGDYLAYNVGQCGEM-YEHSFSNVSGFFTFSYVRTSSGNVF 682
           ||  :  | |:|::    |:|| | ::       : | ||:||:  :   :|   |:  |:
Db     235 LLLEDSYVNLDEINACFSRDGDKLTFSFASESTLNYCHSYSNILKYVCKTYFPASNREVY 294

Qy     683 KLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLD 742
             |:
Db     295 MKEF---------------------------------------------------------- 298

Qy     743 SDFVDRIYSY--ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGL 800
           | |: ::    : | ||                  :    | |
Db     299 --LVTRVNTWFCKFSRIDTF------------------LLYKGVAH------------- 324

Qy     801 VVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNF 860
                     ||: |  |  | | :       : :|:  | |  : :   : |:  |
Db     325 -------------KSVDSEQFYTAMEDAWH-----YKKTLAMCNSERILLEDSSSVNYWF 366

Qy     861 NVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKN---SRLR 917
             :  :: :|                          |||| ||  |  |    |:
Db     367 PKMRDMVIVP---------------------------LFDISLETSKRTRKEVLVSKDF 398

Qy     918 VFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVS 977
           ||:         |:  ::|   | |  |  :| :|:    | |       :: |  |
Db     399 VFT---------VLNHIRTYQAKAL--TYANVLSFVESIRSRV------IINGVTARSEW 441

Qy     978 SVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETS 1037
           | | |:                |    ||                  ::: : :|:
Db     442 DVDKSLLQ--------------SLSMTFY------------------LHTKLAVLKDD 467

Qy    1038 RVAVSKVVAGTK----EFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAA 1093
             : :||   |:|       | |:||                                .
Db     468 -LLISKFSLGSKTVCQHVWDEISL---------------------------------- 490

Qy    1094 NARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVP 1153

Db     491 ---------------------------------------------------------- 490

Qy    1154 LALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAV 1213
                                                                | | : :|
Db     491 -----------------------------------------------------AFGNAFPSV 499

Qy    1214 PEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDY 1273
           | :   | :         | :| ||      :| | ||
```

```
Db    500 KERLLNRKLIR------VAGDALEIR------VPDL-YV---------------------- 525

Qy   1274 DEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQY 1333
Db    526 --------TFHDRL-----------VTEYKASVDMPALDIRKKMEET---------EVMY 557

Qy   1334 TYVEEEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGK 1393
Db    558 NALSE-----------------------LSVLRESDKFDVDVFSQM-----CKSLEVD-- 587

Qy   1394 GLLPESVVGEAPTQERG------RAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQ 1447
Db    588 PMTAAKVIVADMSNESGLTLTFERPTEANVA-LALQDQEKASEGALVVTSREVEEPSMKG 646

Qy   1448 SEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQ 1507
Db    647 S------MARGELQLAGLAGDHP------------------------ESSYSKNEEIES 675

Qy   1508 PKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFS 1567
Db    676 LEQFHMATADSLIRKQMSSI----------VYTGPIKVQQMKN---FIDSLVASLSA--- 719

Qy   1568 KLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYP 1627
Db    720 -------------------AVSNLVKILKDTAAID--LETRQKFG-------------- 743

Qy   1628 SDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRIS 1687
Db    744 --------------------------VLDVASRKWLIKP------------------- 756

Qy   1688 LTFRATRRVFGVGRMLQLAGGVSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGRE 1747
Db    757 ------------------------------------------TAKSHAWGVVETHARK 772

Qy   1748 VKGRSTYSI----WCEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYL 1803
Db    773 ------YHVALLEYDEQGVVTCDDWRR--------------------VAVSSESVV--- 802

Qy   1804 KYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGL 1863
Db    803 -YSDMAKLRTLRRLLRNGEPHVSSAKVVL---------------------------- 830

Qy   1864 RVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVG 1923
Db    831 --------------------VDGV---------------------------------- 834

Qy   1924 ESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGSSEDINMAVKK------- 1976
Db    835 --------------PGCGK--------------TKEILSRVNFDEDLILVPGKQAAEMIR 866

Qy   1977 RDPNLEGLNSAT-----TVNSRVVNF--IVRGMYKRVLVDEVHMMHQGLLQLGVFATGAS 2029
Db    867 RRANSSGIIVATKDNVKTVDSFMMNFGKSTRCQFKRLFIDEGLMLHTGCVNFLVAMSLCE 926

Qy   2030 EGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSS------M 2083
Db    927 IAYVYGDTQQIPYINRVSGFPYPAHFAKLEVDEVETRRTTLRCPADVTHYLNRRYEGFVM 986

Qy   2084 TVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLK 2143
Db    987 STSSVKKSVSQEMVGG--AAVINPISKPSHGK-----------ILTFTQ---SD-KEALL 1029

Qy   2144 GKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFT-KQPHILVGLSRHTRSLVYAA--- 2199
Db   1030 SRG-YSDVHTVHEVQGETYSDVSLVRLTPTPVSIIAGDSPHVLVALSRHTCSLKYYTVVM 1088

Qy   2200 --LSSKLDD--KVGTYISD 2214
Db   1089 DPLVSIIRDLEKLSSYLSD 1107
```

Search completed: January 31, 2001, 15:34:29

Job time: 404 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2001, 15:32:37 ; Search time 102.04 Seconds
(without alignments)
2569.523 Million cell updates/sec

Title: US-09-301-906-15
Perfect score: 11520
Sequence: 1 MDYIRPLRVFSFPHVNNTLE.....QSVSDALLHTFAPAGCFRGI 2237

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3644	31.6	701	12	O71188	O71188 grapevine l
2	803.5	7.0	398	12	Q9Q6Q4	Q9q6q4 grapevine l
3	794.5	6.9	2473	12	O71209	O71209 grapevine l
4	741	6.4	3090	12	Q9Q703	Q9q703 beet yellow
5	736	6.4	3070	12	O89906	O89906 beet yellow
6	734	6.4	3124	12	Q66237	Q66237 citrus tris
7	721	6.3	2301	12	O40959	O40959 little cher
8	721	6.3	3074	12	Q08534	Q08534 sugar beet
9	706.5	6.1	3115	12	Q9WID7	Q9wid7 citrus tris
10	689.5	6.0	3115	12	Q9IFX0	Q9ifx0 citrus tris
11	678.5	5.9	3115	12	O10467	O10467 citrus tris
12	661.5	5.7	3107	12	P87587	P87587 citrus tris
13	643	5.6	1873	12	Q83044	Q83044 lettuce inf
14	483.5	4.2	1334	12	Q9JGK3	Q9jgk3 soil-borne
15	466	4.0	1350	12	Q9WIE3	Q9wie3 chinese whe
16	466	4.0	1858	12	Q9WJD7	Q9wjd7 chinese whe
17	453.5	3.9	1345	12	Q9QBU8	Q9qbu8 oat golden
18	453.5	3.9	1853	12	Q9IWA9	Q9iwa9 oat golden
19	450.5	3.9	1270	12	O93163	O93163 beet soil-b
20	450.5	3.9	1778	12	O56790	O56790 beet soil-b
21	448	3.9	1277	12	Q9JGJ7	Q9jgj7 sorghum chl
22	447.5	3.9	1320	12	Q06359	Q06359 soil-borne
23	447.5	3.9	1828	12	Q89249	Q89249 soil-borne
24	437.5	3.8	1308	12	Q9Q9R0	Q9q9r0 soil-borne
25	437.5	3.8	1816	12	Q9JAD4	Q9jad4 soil-borne
26	436.5	3.8	1306	12	Q9WK10	Q9wk10 broad bean
27	436.5	3.8	1820	12	Q9YPH5	Q9yph5 broad bean
28	427.5	3.7	1308	12	Q9Q9Q6	Q9q9q6 soil-borne
29	427.5	3.7	1816	12	Q9JAD2	Q9jad2 soil-borne
30	423.5	3.7	1308	12	Q9QCE8	Q9qce8 european wh
31	423.5	3.7	1816	12	Q9IWB1	Q9iwb1 european wh
32	422.5	3.7	1308	12	Q9Q9Q8	Q9q9q8 soil-borne
33	422.5	3.7	1816	12	Q9JAD3	Q9jad3 soil-borne
34	413	3.6	1302	12	Q9QBU0	Q9qbu0 potato mop-
35	391	3.4	1707	12	Q96806	Q96806 grapevine t
36	389	3.4	1707	12	Q67704	Q67704 grapevine v
37	386.5	3.4	710	12	Q65853	Q65853 beet yellow
38	373	3.2	1310	12	Q9WKC2	Q9wkc2 beet virus
39	373	3.2	1818	12	O90699	O90699 beet virus
40	359	3.1	2038	12	Q9IH80	Q9ih80 cherry necr
41	344	3.0	1180	12	Q84919	Q84919 pepper ring
42	344	3.0	1698	12	Q84918	Q84918 pepper ring
43	338.5	2.9	1188	12	P90434	P90434 tobacco rat
44	337	2.9	1130	12	P89202	P89202 sunnhemp mo
45	336.5	2.9	1707	12	O56250	O56250 tobacco rat

ALIGNMENTS

RESULT 1
O71188
ID O71188 PRELIMINARY; PRT; 701 AA.
AC O71188;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 77.0 KDA PROTEIN (FRAGMENT).
OS grapevine leafroll-associated virus 3.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=55951;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NY1;
RX MEDLINE=98264508; PubMed=9603346;
RA Ling K.-S., Zhu H.-Y., Drong R.F., Slightom J.L., McFerson J.R.,
RA Gonsalves D.;
RT "Nucleotide sequence of the 3'-terminal two-thirds of the grapevine
RT leafroll-associated virus-3 genome reveals a typical monopartite
RT closterovirus.";
RL J. Gen. Virol. 79:1299-1307(1998).
DR EMBL; AF037268; AAC40717.1; -.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicase1; 1.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 701 AA; 77008 MW; 772A764D2CA2A2DC CRC64;

Query Match 31.6%; Score 3644; DB 12; Length 701;
Best Local Similarity 100.0%; Pred. No. 3.8e-248;
Matches 701; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1537 DVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAI 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 DVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAI 60

Qy   1597 KYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVI 1656
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVI 120
```

```
Qy   1657 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPG 1716
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 NVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPG 180

Qy   1717 VPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVM 1776
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 VPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVM 240

Qy   1777 ALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPN 1836
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPN 300

Qy   1837 QVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLF 1896
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 QVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLF 360

Qy   1897 VCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNS 1956
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 VCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNS 420

Qy   1957 TATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQ 2016
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQ 480

Qy   2017 GLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDV 2076
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDV 540

Qy   2077 CYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKS 2136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CYLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKS 600

Qy   2137 DMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLV 2196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 DMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLV 660

Qy   2197 YAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAGCFRGI 2237
          |||||||||||||||||||||||||||||||||||||||||
Db    661 YAALSSKLDDKVGTYISDASPQSVSDALLHTFAPAGCFRGI 701
```

```
RESULT   2
Q9Q6Q4
ID   Q9Q6Q4      PRELIMINARY;      PRT;   398 AA.
AC   Q9Q6Q4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   HELICASE (FRAGMENT).
GN   HEL.
OS   grapevine leafroll-associated virus 1.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=47985;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fazeli C.F., Rezaian M.A.;
RT   "Nucleotide sequence and organization of ten open reading frames of
RT   the grapevine leafroll-associated virus 1 genome and identification of
RT   three subgenomic RNAs.";
RL   J. Gen. Virol. 0:0-0(2000).
DR   EMBL; AF195822; AAF22737.1; -.
DR   INTERPRO; IPR000606; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
KW   Helicase.
FT   NON_TER       1      1
SQ   SEQUENCE   398 AA;  44426 MW;  A908109D49C22A23 CRC64;
```

```
  Query Match             7.0%;  Score 803.5;  DB 12;  Length 398;
  Best Local Similarity   44.2%;  Pred. No. 3.8e-48;
  Matches  174;  Conservative   59;  Mismatches  156;  Indels    5;  Gaps    2;
```

```
Qy   1847 VYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHD 1906
          || :   :: :: :||  : || || :   | |   ::: ||::  | : ||||::||||
Db      3 VYDANFKVVANAPEDGTTIFDLEYVFLVSTGTFVPRRNLQAVLSRQDAVLVCDELLVFHD 62

Qy   1907 AINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITANVGS 1966
          |:|| | :::|:  |||| |       |:||||| || ||||:           || ||
Db     63 AMNLRGCVRLAKRAMVGEYMKDVRISAVNSPPGGGNTTRLVDEYFGRKKRAKIAAANTGS 122

Qy   1967 SEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFAT 2026
            ||| |:: |:   |    | | || |:|   |       |:|||:|:|:|: || | :
Db    123 VADINAAIRAREGKKEPDLVAKTANSWVINSHPRPNSHVGLIDEVYMLHKGMFQLTVVSM 182

Qy   2027 GASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVR 2086
          |  | : :|| |||||||||| |          :| : ||  ||||| |||||||||
Db    183 GVKEVIAYGDKNQIPFINREKTFVTPNEAVEFAEEQIDYTDISYRCPADVCYVLSSMTDA 242

Qy   2087 GTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKG-- 2144
            :| ||  |    |   :||  | || | :|     ||||| ||| ||::|::::
Db    243 RGKKMYPNGVFPNGDVRPLRSFEKVPIATPEDALLYEADVYLTMTQNEKAEMQKAVAKME 302

Qy   2145 --KGKETP-VMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALS 2201
             ||: | |:| |||||||: :||| | ||||| :|:::|||:| ||||||: || ||
Db    303 VVAGKKRPDVITTHEAQGKTYENVVLVRLKKADDPIFSRKPHIVVALSRHTRSMKYAVLS 362

Qy   2202 SKLDDKVGTYISDASPQSVSDALLHTFAPAGCFR 2235
          ||: | :   |   |   ||| ||       ||
Db    363 SKMTDTISKLIDGTSAGKVSDVLLQQLQRNDRFR 396
```

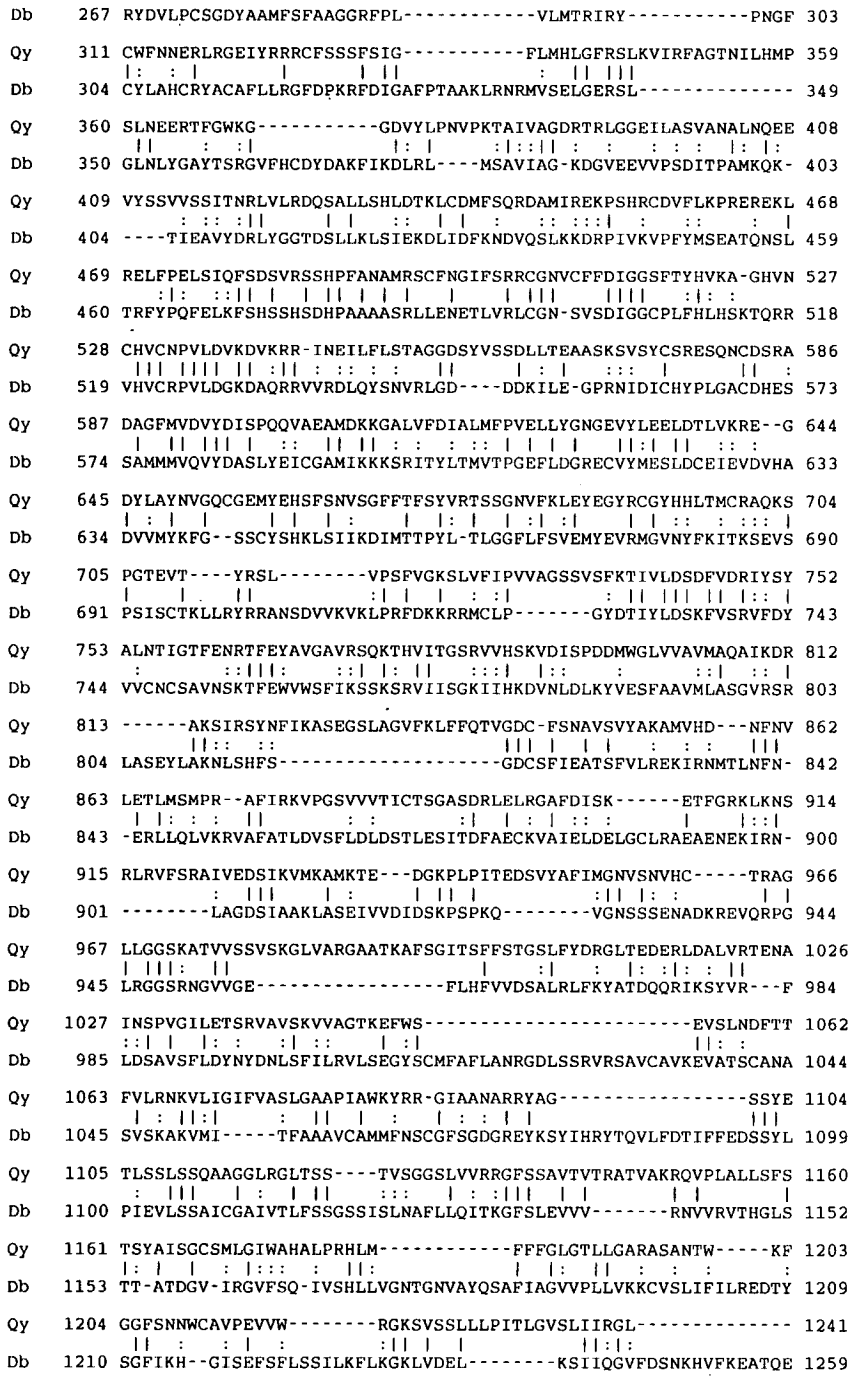
```
RESULT   3
O71209
ID   O71209      PRELIMINARY;      PRT;  2473 AA.
AC   O71209;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   METHYLTRANSFERASE/HELICASE POLYPROTEIN (FRAGMENT).
OS   grapevine leafroll-associated virus 2.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=64003;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98264507; PubMed=9603345;
RA   Zhu H.Y., Ling K.S., Goszczynski D.E., McFerson J.R., Gonsalves D.;
RT   "Nucleotide sequence and genome organization of grapevine leafroll-
RT   associated virus-2 are similar to beet yellows virus, the
RT   closterovirus type member.";
RL   J. Gen. Virol. 79:1289-1298(1998).
DR   EMBL; AF039204; AAC40855.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; Methyltransferase; Helicase; Polyprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   2473 AA;  277096 MW;  F8DEEBA6A58ED425 CRC64;
```

```
  Query Match             6.9%;  Score 794.5;  DB 12;  Length 2473;
  Best Local Similarity   21.6%;  Pred. No. 6.5e-46;
  Matches  538;  Conservative  342;  Mismatches  913;  Indels  699;  Gaps  113;
```

```
Qy    152 QDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEP-AQSVPQQSL---------EEKAAL 201
          |   | ||: :|  |:  | |  |  |:::   |     ||                | :  :
Db    158 QSSVKTKREDAHRTVEERAAGGSVEQPRQKRIDEKGCGRVPSGGFSHLLVGNLNEVRRKV 217

Qy    202 TFGLFFSKGGGDESDAVILRKGKLFNRALNVPIDVKNTFVWAK--------IWDEASRRR 253
            ||   : |||            |:|: :         || :        ::  :
Db    218 AAGLLRFRVGGDMD----------FHRSFSTQAG-HRLLVWRRSSRSVCLELYSPSKNFL 266

Qy    254 GYFYVK---DRAVKFFPIVRGRATIEDFIVNTAPGCDVALPRIELWSMRERAFVCTTKGW 310
            |  :    | |  |     ||  :            | : ||               |:
```


Db 267 RYDVLPCSGDYAAMFSFAAGGRFPL----------VLMTRIRY----------PNGF 303
Qy 311 CWFNNERLRGEIYRRRCFSSSFSIG----------FLMHLGFRSLKVIRFAGTNILHMP 359
Db 304 CYLAHCRYACAFLLRGFDPKRFDIGAFPTAAKLRNRMVSELGERSL---------- 349
Qy 360 SLNEERTFGWKG----------GDVYLPNVPKTAIVAGDRTRLGGEILASVANALNQEE 408
Db 350 GLNLYGAYTSRGVFHCDYDAKFIKDLRL----MSAVIAG-KDGVEEVVPSDITPAMKQK- 403
Qy 409 VYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPREREKL 468
Db 404 ----TIEAVYDRLYGGTDSLLKLSIEKDLIDFKNDVQSLKKDRPIVKVPFYMSEATQNSL 459
Qy 469 RELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKA-GHVN 527
Db 460 TRFYPQFELKFSHSSHSDHPAAAASRLLENETLVRLCGN-SVSDIGGCPLFHLHSKTQRR 518
Qy 528 CHVCNPVLDVKDVKRR-INEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCDSRA 586
Db 519 VHVCRPVLDGKDAQRRVVRDLQYSNVRLGD----DDKILE-GPRNIDICHYPLGACDHES 573
Qy 587 DAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEELDTLVKRE--G 644
Db 574 SAMMMVQVYDASLYEICGAMIKKKSRITYLTMVTPGEFLDGRECVYMESLDCEIEVDVHA 633
Qy 645 DYLAYNVGQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKS 704
Db 634 DVVMYKFG--SSCYSHKLSIIKDIMTTPYL-TLGGFLFSVEMYEVRMGVNYFKITKSEVS 690
Qy 705 PGTEVT----YRSL--------VPSFVGKSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSY 752
Db 691 PSISCTKLLRYRRANSDVVKVKLPRFDKKRRMCLP-------GYDTIYLDSKFVSRVFDY 743
Qy 753 ALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDR 812
Db 744 VVCNCSAVNSKTFEWVWSFIKSSKSRVIISGKIIHKDVNLDLKYVESFAAVMLASGVRSR 803
Qy 813 ------AKSIRSYNFIKASEGSLAGVFKLFFQTVGDC-FSNAVSVYAKAMVHD---NFNV 862
Db 804 LASEYLAKNLSHFS----------GDCSFIEATSFVLREKIRNMTLNFN- 842
Qy 863 LETLMSMPR--AFIRKVPGSVVVTICTSGASDRLELRGAFDISK------ETFGRKLKNS 914
Db 843 -ERLLQLVKRVAFATLDVSFLDLDSTLESITDFAECKVAIELDELGCLRAEAENEKIRN- 900
Qy 915 RLRVFSRAIVEDSIKVMKAMKTE---DGKPLPITEDSVYAFIMGNVSNVHC-----TRAG 966
Db 901 --------LAGDSIAAKLASEIVVDIDSKPSPKQ--------VGNSSSENADKREVQRPG 944
Qy 967 LLGGSKATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENA 1026
Db 945 LRGGSRNGVVGE----------FLHFVVDSALRLFKYATDQQRIKSYVR---F 984
Qy 1027 INSPVGILETSRVAVSKVVAGTKEFWS----------EVSLNDFTT 1062
Db 985 LDSAVSFLDYNYDNLSFILRVLSEGYSCMFAFLANRGDLSSRVRSAVCAVKEVATSCANA 1044
Qy 1063 FVLRNKVLIGIFVASLGAAPIAWKYRR-GIAANARRYAG----------SSYE 1104
Db 1045 SVSKAKVMI-----TFAAAVCAMMFNSCGFSGDGREYKSYIHRYTQVLFDTIFFEDSSYL 1099
Qy 1105 TLSSLSSQAAGGLRGLTSS----TVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFS 1160
Db 1100 PIEVLSSAICGAIVTLFSSGSSISLNAFLLQITKGFSLEVVV-------RNVVRVTHGLS 1152
Qy 1161 TSYAISGCSMLGIWAHALPRHLM----------FFFGLGTLLGARASANTW-----KF 1203
Db 1153 TT-ATDGV-IRGVFSQ-IVSHLLVGNTGNVAYQSAFIAGVVPLLVKKCVSLIFILREDTY 1209
Qy 1204 GGFSNNWCAVPEVVW--------RGKSVSSLLLPITLGVSLIIRGL---------- 1241
Db 1210 SGFIKH--GISEFSFLSSILKFLKGKLVDEL--------KSIIQGVFDSNKHVFKEATQE 1259
Qy 1242 -LNDTIPQ--LAYVPPVE---GRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDG- 1294
Db 1260 AIRTTVMQVPVAVVDALKSAAGKIYNNFTSR--RTFGKDEGSSSDGACEEYFSCDEGEGP 1317
Qy 1295 STSSVSSYD---------VVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAA 1345
Db 1318 GLKGGSSYGFSILAFFSRIMWGARRLIVKVKHECFGK---------LFEFL------SLK 1362
Qy 1346 VAERQGDPSGSGTADAMAFVESVKKGVDD-----VFHQQSSGETAREVEVDGKGLLPESV 1400
Db 1363 LHEFRTRVFGKNRTDV---------GVYDFLPTGIVETLSSIEECDQIE----ELLGDDL 1409
Qy 1401 VGEAPTQERGRAADGNTAQTAVNEGD-REPVQSSLVSSPQADIPKVTQS---------- 1448
Db 1410 KG----------DKDASLTDMNYFEFSEDFLASIEEPPFAGLRGGSKNIAILAILEYAH 1458
Qy 1449 ---EVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQV 1505
Db 1459 NLFRIVASKCSKRPLFLAFAELSSALIEK--FKEVFPR------KSQLVAIVREYTQ-RF 1509
Qy 1506 EQPKQRSLTINEGKAGKQLC------MFRTCSCGVQLDVYNE-------ATIATRFSNAF 1552
Db 1510 LRSRMRALGLNNEFVVKSFADLLPALMKRKVSGSFLASVYRPLRGFSYMCVSAERREKFF 1569
Qy 1553 TFVDNLKGRSAVFFSK----------LGEGYT---------- 1574
Db 1570 ALV-CLIGLSLPFFVRIVGAKACEELVSSARRFYERIKIFLRQKYVSLSNFFCHLFSSDV 1628
Qy 1575 --------YNGGSH----------VSSGWPRALEDILTAIKYPSVFDHC--LVQK 1609
Db 1629 DDSSASAGLKGGASRMTLFHLLVRLASALLSLGW----EGLKLLLSHHNLLFLCFALVDD 1684
Qy 1610 ----YKMGGGVPFHADDEECYPSDNPILTV----------NLVGKANFSTKCRK 1649
Db 1685 VNVLIKVLGGLSFFV---------QPIFSLFAAMLLQPDRFVEYSEKLVTAFEFFLKCSP 1735
Qy 1650 GGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVF------GVGRML 1703
Db 1736 RAPALL----------KGF-------FECVANSTVSKTVRRLLRCFVKMLKLRKGRGL 1776
Qy 1704 QLAG-GVSDEKS-PGVPNQQPQSQGATRTITPKSGGKALS----------EGSGRE 1747
Db 1777 RADGRGLHRQKAVPVIPSNRVVTDGVERLSVKMQGVEALRTELRILEDLDSAVIEKLNRR 1836
Qy 1748 VKGRSTYSIWCEQDYVRK----------CEWLRADNPVMALE----------PDYTPM 1785
Db 1837 -RNRDTN----DDEFTRPAHEQMQEVTTFCS--KANSAGLALERAVLVEDAIKSEKLSKT 1889
Qy 1786 TFEVVKTG--TSEDAVV-----EYLKYLAIGIER-----TYR----------ALLMARNI 1823
Db 1890 VNEMVRKGSTTSEEVAVALSDDEAVEEISVADERDDSPKTVRISEYLNRLNSSFEFPKPI 1949
Qy 1824 AVTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHST----QDGLRV----------RDL 1868
Db 1950 VVDDNKDTGGLTNAVRE----FYYMQELALFEIHSKLCTYYDQLRIVNFDRSVAPCSEDA 2005
Qy 1869 PYVLIAEKGIFTKGKDVDA-VVALGDNLFVCDDIL----------VFHDAINLIGALKV 1916
Db 2006 QLYVRKNGSTIVQGKEVRLHIKDFHDHDFLFDGKISINKRRRGGNVLYHD--NL--AFLA 2061
Qy 1917 ARCGMVGESF--------KSFEYKCYNAPPGGGKTTTLVDEFV----KSPNSTATITANV 1964
Db 2062 SNLFLAGYPFSRSFVFTNSSVDILLYEAPPGGGKTTTLIDSFLKVFKKGEVSTMILTANK 2121
Qy 1965 GSSEDINMAVKKRDPNLE---------GLNSATTVNSRVVNFIVRGMYKRVL-VDEVHMM 2014
Db 2122 SSQVEILKKVEKEVSNIECQKRKDKRSPKKSIYTIDAYLMHH--RGCDADVLFIDECFMV 2179
Qy 2015 HQGLLQLGVFATGASEGLFFGDINQIPFINR---EKVFRMDCAVFVPKKESVVYTSKSYR 2071
Db 2180 HAGSVLACIEFTRCHKVMIFGDSRQIHYIERNELDKCLYGDLDRFV-DLQCRVYGNISYR 2238

```
Qy   2072 CPLDVCYLLSSM------TVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINAD 2125
          || |||  ||::      ||:|  :        ||      |:    | : ||:
Db   2239 CPWDVCAWLSTVYGNLIATVKGESE--------GKS-----SMRINEINSVDDLVPDVGS 2285

Qy   2126 VYLCMTQLEKSDMKRSLKGKG-KETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHI 2184
          :||| | || :: :    ||  :  |:|||||||:|:: | | | |  :|  |    ||
Db   2286 TFLCMLQSEKLEISKHFIRKGLTKLNVLTVHEAQGETYARVNLVRLKFQEDEPFKSIRHI 2345

Qy   2185 LVGLSRHTRSLVYAALSSKLDDKVGTYISDAS 2216
           | |||| || |  |:::  |     |  |:
Db   2346 TVALSRHTDSLTYNVLAARRGDATCDAIQKAA 2377


RESULT  4
Q9Q703
ID   Q9Q703      PRELIMINARY;      PRT;  3090 AA.
AC   Q9Q703;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P348.
OS   Beet yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID-12161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BYV-4;
RX   MEDLINE=20079557; PubMed=10611288;
RA   Peremyslov V.V., Hagiwara Y., Dolja V.V.;
RT   "HSP70 homolog functions in cell-to-cell movement of a plant virus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:14771-14776(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BYV-4;
RA   Peremyslov V.V., Hagiwara Y., Alzhanova D., Dolja V.V.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190581; AAF14300.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   3090 AA;  347990 MW;  09C779A57DE874B9 CRC64;


  Query Match             6.4%;  Score 741;  DB 12;  Length 3090;
  Best Local Similarity   22.6%;  Pred. No. 6.1e-42;
  Matches  498;  Conservative  272;  Mismatches  776;  Indels  654;  Gaps  100;

Qy    408 EVYSSV----VSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPR 463
          |: ||:    : |:  |: :   : |   ::  | |   :  :: :|| |     ::
Db    596 EITSSLKHKAIESVYERVSIHKDNLLARSVEKDLIDFKDEIKSLSKEKRSVTVPFYMGEA 655

Qy    464 EREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKA 523
          :  |   :|: :: |: || | || |   |   |   :    :  | ||||   :|:|
Db    656 VQSGLTRAYPQFNLSFTHSVYSDHPAAAGSRLLENETLASMAKS-SFSDIGGCPLFHIKR 714

Qy    524 GHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCD 583
          |  : ||| |: |:|| :||::    |   |    :| : | || :: || |     ||:
Db    715 GSTDYHVCRPIYDMKDAQRRVSR--ELQARGLVENLSREQLVEAQAR-VSVCPHTLGNCN 771

Qy    584 SRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEEL--DTLVK 641
          ::|   || ||| |  ::| ||  | : |  : :: | |||       :: |  | :|
Db    772 VKSDVLIMVQVYDASLNEIASAMVLKESKVAYLTMVTPGELLDEREAFAIDALGCDVVVD 831

Qy    642 REGDYLAYNVGQCGEMYEHSFSNVSGF-----FTFSYVRTSSGNVFKLEYEGYRCGYHHL 696
             | : |  |     | |  ||:        |||      |||:| :|    | | ::
Db    832 TRRDMVQYKFG--SSCYCHKLSNIKNIMLTPAFTF------SGNLFSVEMYENRMGVNYY 883

Qy    697 TMCRAQKSPG------------TEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDS 743
          : |:  ||              |||    | | |      |:         :  | ||:

Db    884 KITRSAYSPEIRGVKTLRYRRACTEVVQVKL-PRFDKTLKTFLS-------GYDYIYLDA 935

Qy    744 DFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVA 803
          || |:: | ::      ::|||:   ::| |: |:  :|:|  | |
Db    936 KFVSRVFDYVVSNCSVVNSKTFEWVWSYIKSSKSRVVISGKVIHRDVHIDLKHSECFAAV 995

Qy    804 VMAQAIKDR------AKSIRSYNFIKASEGSLAGVFKLF----FQTVGDCFSNAVSVYAK 853
          ::|  :: |      ||::  |    :   ::  :|: :    :  :   |   :
Db    996 MLAVGVRSRTTTEFLAKNLNYYTGDASCFETIRFLFREWSRRAYAEINRSFRKLMKSILS 1055

Qy    854 AMVHDNF----NVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGR 909
          | :   |    | |: |:      :|       |:|  :|  |
Db   1056 AGLDYEFLDLDNSLQHLLEYSEVEVR-------VSIAQNGEVD---------------- 1091

Qy    910 KLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHC-TRAGLL 968
          |   || :  | |              | |       |:   : | :|:|    | ||
Db   1092 --CNEENRVLTEIIAE----------AADRK-------SIAQGLSGALSSVPTQPRGGLR 1132

Qy    969 GGSKAT-------VVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALV 1021
          |||: :       :|  |      : | | :    :   ||   :|    :   |     |
Db   1133 GGSRRSGVSFLYNLVEEVGNLFFSVGDAVRFLVKVFKTFSDSPIFRVVRMFLD-----LA 1187

Qy   1022 RTENAINSPVGILETSRVAVS--------KVVAGTKEFWSEVSLNDFTTFVLRNKVLIGI 1073
             :  | | :    | |||         : |: : : :   ::  |  |:    :
Db   1188 EAASPFVSVVSLCAWLREAVSAFSSWVADRTVSESVKTFVNRTVKRFLNFMSAKTLTKKF 1247

Qy   1074 FVASLGAAPIA--------------WK-YRRGIAANARRYAGSSYETLSSLSSQAAGGLR 1118
          |   | |: :|                |: :   | :::  |  |: |  ||:
Db   1248 FRFFLSASALAKTVVRKAKVILEAYWEVWFESILSDSGEY--SAVEFCSSVV-------- 1297

Qy   1119 GLTSSTVSGGSLVVRRGFSSAVTVTR-----ATVAKRQVPLALLSFSTSYAISG------ 1167
          :|  | ||  |    ||| :  :|       ||    :| |  :|   | | |
Db   1298 -ITLLTNSGRLL---PGFSPSAVITEVLLDLATKISIEVLLKQISPVDSTASSALYRRVL 1353

Qy   1168 ----------------------CSMLGIW-----AHALPRHLMFF-----FGLGTLLGAR 1195
                                 |  | ::     |  :|  :  :     :|:  |
Db   1354 SEILSNFRTMGEHGIFTKVFLLCGFLPVFVRKCVALCVPGDMATYARFLEYGVDDLFFLG 1413

Qy   1196 ASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPIT-------LGV-SLIIRGLLN---- 1243
          | |: |     |:  |          | |::| ::        ||  | ||:  ||
Db   1414 RSVNSIK------NYLCVVAAGLVDSIVDSVVLKLSGVAKERVLGFKSKIIKNFLNVFRK 1467

Qy   1244 ----------------------DTIPQL---------------------------------- 1249
                                  :: | |
Db   1468 AKVVTRTSSSTDLSEDEYFSCDESKPGLRGGSSRFTLSRLLDIFFNFLKSSKLVIENACF 1527

Qy   1250 -AYVPPVEGRNVY--------DETLRYYR---DFDY------DEGAGPSGTQHEAVPGDD 1291
           ||        :|        :|  |  |   ||||      ||         :    ||
Db   1528 SAYERIERNMKLYFFPLNSSEEEARRLIRCAGDFDYLSDSAFDEDEMLRQAFEQYYSSDD 1587

Qy   1292 N----DGSTSSVSSYDVVTNVRDVGISTNG-----EVTGEEETHSPRSVQ-YTYVEEEVA 1341
             ||  : : ||  |:  | :    ||     :|:   :    | ::   :|:  ::
Db   1588 ESVTYDGKPTVLRSYLNVSR-RFLETFCNGPKFFVKVSNYFKALYSRLLRVLPWVDRNLS 1646

Qy   1342 PSAAVAERQGDPSGSGTADAMAFV--------ESVKK----------------GVDDVFH 1377
           |       |   |:  |     |:         | | :                  |:  :|:
Db   1647 DS------PGLKGGNEKALLAKFLKTCVITACECVSQICCLRLIRLCWGTPACGLVRLFY 1700

Qy   1378 QQSSGETAREVEVDGKGLLPESVVGEAP------------------TQERGRAADGNTA 1418
             ||       |    : |  |  |                        |  :|  |:|
Db   1701 ITYSGTRVLSRVVVAVAVCPLLVRNELDGLSDGLTNMGVSVFRRLFVALRRALSAYSNSA 1760

Qy   1419 -----------------QTAV---NEGDREPVQSSLVSSPQADIPK----VTQSEVHAQK 1454
                            ||   ||   ||:      ||: ||       : ||  :
Db   1761 LRRKIFEFIFGNIHHPFDVAVIETNEVAPEPL------SPEVDIDVDCDFGSDSESVSSD 1814

Qy   1455 EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDK--GKA----------VAHVAEKKQ 1502
          ||   :|   : | :     : : :  |  ||:: |  |:           ||:  |::
Db   1815 EV-ASIPRPGLHGGS-----RRSSNFLTSLVKVVFKLAGRIPRLLFRLRNFVAYFVERR- 1867

Qy   1503 VQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRS 1562
                      | |:|  |                  :|  |       ||    |
Db   1868 ----------------LASKRLKTF--------------IGLARLF-------DNFSLTS 1890
```

```
Qy   1563 AVFF-----SKLG-----EGYTYNGGS-----HVSSGWPRA----LEDILTAIKYPSVFD 1603
Db   1891 VVYLLQEYDSVLNAFIDVELVLLNSGSVNVLPLVS--WVRGSLTKLAEVIVGSGFASFLG 1948

Qy   1604 H--CLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASG 1661
Db   1949 RMCCRVSDW-------------CSSSSN--------AGCNFMSPVRTKGK-FVPPSSSG 1985

Qy   1662 DYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEK-------- 1713
Db   1986 STASM---YER--LEALESDIREHVLSTCR-------VG---------SDEEEERPKEVT 2024

Qy   1714 SPGVPNQQPQSQGATRTITPKSGG--------------------KALSEGSGREVKGRS 1752
Db   2025 EPGIEHTSEDVVPIRSHSQPLSGGECSYSEDREENERANLLPHVSKIVSERRGLETARRN 2084

Qy   1753 TYSIWCEQDYVRKCEWLRADN-------PVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805
Db   2085 ------KRTLRGVSEFLNAINTSNEQPRPIIV---DHSPES----RALTNSVREFYYLQE 2131

Qy   1806 LAI-----GIERTYRALLMAR-------------NIAVTTAEGVLKVPNQVYESLPGFHV 1847
Db   2132 LALFELSCKLREYYDQLKVANFNRQECLCDKDEDMFVLRAGQGVVSGRN---SRLPLKHF 2188

Qy   1848 YKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDA 1907
Db   2189 --KGHEFCFRS--GGL----VPY---------------------DGTSRVDTI--FHTQ 2216

Qy   1908 INLIGALKVARCGMVGESFKSFEYK-------CYNAPPGGGKTTTLVDEFVK--SPNSTA 1958
Db   2217 TNFVSA-NALLSGYL--SYRTFTFTNLSANVLLYEAPPGGGKTTTLIKVFCETFSKVNSL 2273

Qy   1959 TITANVGSSEDI-----NMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVL-VDEVH 2012
Db   2274 ILTANKSSREEILAKVNRIVLDEGDTPLQTRDRILTIDSYLMN--NRGLTCKVLYLDECF 2331

Qy   2013 MMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAV------FVPKKESVVYT 2066
Db   2332 MVHAGAAVACIEFTKCDSAILFGDSRQIHYIDRNE---LDTAVLSDLNRFV-DDESRVYG 2387

Qy   2067 SKSYRCPLDVCYLLSSMTVRGTEKCYPEKVV------SGKDKPVVRSLSKRPIGTTDDVA 2120
Db   2388 EVSYRCPWDVCAWLSTF--------YPKTVATTNLVSAGQS-----SMQVREIESVDDVE 2434

Qy   2121 EINADVYLCMTQLEKSDM-----KRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADD 2175
Db   2435 YSSEFVYLTMLQSEKKDLLKSFGKRSRSSVEKPT-VLTVHEAQGETYRKVNLVRTKFQED 2493

Qy   2176 SLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDA 2215
Db   2494 DPFRSENHITVALSRHVESLTYSVLSSKRDDAIAQAIVKA 2533


RESULT   5
O89906
ID   O89906      PRELIMINARY;      PRT;  3070 AA.
AC   O89906;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   FUSION PROTEIN.
OS   Beet yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12161;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CALIFORNIAN;
RX   MEDLINE=98285746; PubMed=9621048;
RA   Peremyslov V.V., Hagiwara Y., Dolja V.V.;
RT   "Genes required for replication of the 15.5-kilobase RNA genome of a
RT   plant closterovirus.";
RL   J. Virol. 72:5870-5876(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CALIFORNIAN;
RA   Dolja V., Peremyslov V., Hagivara Y.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF056575; AAC25115.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
FT   CHAIN         1    588       PAPAIN-LIKE LEADER PEPTIDE.
FT   CHAIN       589   2042       PUTATIVE METHYLTRANSFERASE AND RNA
FT                                HELICASE.
FT   CHAIN      2626   3070       RNA-DEPENDENT RNA POLYMERASE.
SQ   SEQUENCE   3070 AA;  345803 MW;  6FC307093F1DF36E CRC64;


  Query Match             6.4%;  Score 736;  DB 12;  Length 3070;
  Best Local Similarity   22.5%;  Pred. No. 1.4e-41;
  Matches  497;  Conservative  273;  Mismatches  770;  Indels  666;  Gaps  100;

Qy    408 EVYSSV----VSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPR 463
Db    596 EITSSLKHKAIESVYERVSIHKDNLLARSVEKDLIDFKDEIKSLSKEKRSVTVPFYMGEA 655

Qy    464 EREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKA 523
Db    656 VQSGLTRAYPQFNLSFTHSVYSDHPAAAGSRLLENETLASMAKS-SFSDIGGCPLFHIKR 714

Qy    524 GHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDSYVSSDLLTEAASKSVSYCSRESQNCD 583
Db    715 GSTDYHVCRPIYDMKDAQRRVSR--ELQARGLVENLSREQLVEAQAR-VSVCPHTLGNCN 771

Qy    584 SRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEEL--DTLVK 641
Db    772 VKSDVLIMVQVYDASLNEIASAMVLKESKVAYLTMVTPGELLDEREAFAIDALGCDVVVD 831

Qy    642 REGDYLAYNVGQCGEMYEHSFSNVSGF-----FTFSYVRTSSGNVFKLEYEGYRCGYHHL 696
Db    832 TRRDMVQYKFG--SSCYCHKLSNIKNIMLTPAFTF------SGNLFSVEMYENRMGVNYY 883

Qy    697 TMCRAQKSPG-------------TEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDS 743
Db    884 KITRSAYSPEIRGVKTLRYRRACTEVVQVKL-PRFDKTLKTFLS-------GYDYIYLDA 935

Qy    744 DFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVA 803
Db    936 KFVSRVFDYVVSNCSVVNSKTFEWVWSYIKSSKSRVVISGKVIHRDVHIDLKHSECFAAV 995

Qy    804 VMAQAIKDR------AKSIRSYNFIKASEGSLAGVFKLF----FQTVGDCFSNAVSVYAK 853
Db    996 MLAVGVRSRTTTEFLAKNLNYYTGDASCFETIRFLFREWSRRAYAEINRSFRKLMKSILS 1055

Qy    854 AMVHDNF----NVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGR 909
Db   1056 AGLDYEFLDLDNSLQHLLEYSEVEVR-------VSIAQNGEVD---------------- 1091

Qy    910 KLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHC-TRAGLL 968
Db   1092 --CNEENRVLTEIIAE----------AADRK-------SIAQGLSGALSSVPTQPRGGLR 1132

Qy    969 GGSKAT-------VVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALV 1021
Db   1133 GGSRRSGVSFLYNLVEEVGNLFFSVGDAVRFLVKVFKTFSDSPIFRVVRMFLD-----LA 1187

Qy   1022 RTENAINSPVGILETSRVAVS--------KVVAGTKEFWSEVSLNDFTTFVLRNKVLIGI 1073
Db   1188 EAASPFVSVVSLCAWLREAVSAFSSWVADRTVSESVKTFVNRTVKRFLNFMSAKTLTKKF 1247

Qy   1074 FVASLGAAPIA--------------WK-YRRGIAANARRYAGSSYETLSSLSSQAAGGLR 1118
```

```
Db   1248 FRFFLSASALAKTVVRKAKVILEAYWEVWFESILSDSGEY--SAVEFCSSVV-------- 1297

Qy   1119 GLTSSTVSGGSLVVRRGFSSAVTVTR-----ATVAKRQVPLALLSFSTSYAISG------ 1167

Db   1298 -ITLLTNSGRLL---PGFSPSAVITEVLLDLATKISIEVLLKQISPVDSTASSALYRRVL 1353

Qy   1168 ----------------------CSMLGIW-----AHALPRHLMFF-----FGLGTLLGAR 1195

Db   1354 SEILSNFRTMGEHGIFTKVFLLCGFLPVFVRKCVALCVPGDMATYARFLEYGVDDLFFLG 1413

Qy   1196 ASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPIT-------LGV-SLIIRGLLN---- 1243

Db   1414 RSVNSIK------NYLCVVAAGLVDSIVDSVVLKLSGVAKERVLGFKSKIIKNFLNVFRK 1467

Qy   1244 ----------------------DTIPQL-------------------------------- 1249

Db   1468 AKVVTRTSSSTDLSEDEYFSCDESKPGLRGGSSRFTLSRLLDIFFNFLKSSKLVIENACF 1527

Qy   1250 -AYVPPVEGRNVY--------DETLRYYR---DFDY------DEGAGPSGTQHEAVPGDD 1291

Db   1528 SAYERIERNMKLYFFPLNSSEEEARRLIRCAGDFDYLSDSAFDEDEMLRQAFEQYYSSDD 1587

Qy   1292 N----DGSTSSVSSYDVVTNVRDVGISTNG-----EVTGEEETHSPRSVQ-YTYVEEEVA 1341

Db   1588 ESVTYDGKPTVLRSYLNVSR-RFLETFCNGPKFEVKVSNYFKALYSRLLRVLPWVDRNLS 1646

Qy   1342 PSAAVAERQGDPSGSGTADAMAFV--------ESVKK----------------GVDDVFH 1377

Db   1647 DS------PGLKGGNEKALLAKFLKTCVITACECVSQICCLRLIRLCWGTPACGLVRLFY 1700

Qy   1378 QQSSGETAREVEVDGKGLLPESVVGEAP-------------------TQERGRAADGNTA 1418

Db   1701 ITYSGTRVLSRVVVAVAVCPLLVRNELDGLSDGLTNMGVSVFRRLFVALRRALSAYSNSA 1760

Qy   1419 -----------------QTAV---NEGDREPVQSSLVSSPQADIPK----VTQSEVHAQK 1454

Db   1761 LRRKIFEFIFGNIHHPFDVAVIETNEVAPEPL------SPEVDIDVDCDFGSDSESVSSD 1814

Qy   1455 EVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDK--GKA----------VAHVAEKKQ 1502

Db   1815 EV-ASIPRPGLHGGS-----RRSSNFLTSLVKVVFKLAGRIPRLLFRLRNFVAYFVERR- 1867

Qy   1503 VQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRS 1562

Db   1868 ----------------LASKRLKTF--------------IGLARLF-------DNFSLTS 1890

Qy   1563 AVFF-----SKLG-----EGYTYNGGS-----HVSSGWPRA----LEDILTAIKYPSVFD 1603

Db   1891 VVYLLQEYDSVLNAFIDVELVLLNSGSVNVLPLVS--WVRGSLTKLAEVIVGSGFASFLG 1948

Qy   1604 H--CLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASG 1661

Db   1949 RMCCRVSDW-------------CSSSSN--------AGCNFMSPVRTKGK-FVPPSSSG 1985

Qy   1662 DYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGGVSDEK-------- 1713

Db   1986 STASM---YER--LEALESDIREHVLSTCR-------VG---------SDEEEERPKEVT 2024

Qy   1714 SPGVPNQQPQSQGATRTITPKSGG--------------------KALSEGSGREVKGRS 1752

Db   2025 EPGIEHTSEDVVPIRSHSQPLSGGECSYSEDREENERANLLPHVSKIVSERRGLETARRN 2084

Qy   1753 TYSIWCEQDYVRKCEWLRADN-------PVMALEPDYTPMTFEVVKTGTSEDAVVEYLKY 1805

Db   2085 ------KRTLRGVSEFLNAINTSNEQPRPIIV---DHSPES----RALTNSVREFYYLQE 2131

Qy   1806 LAI-----GIERTYRALLMAR-------------NIAVTTAEGVLKVPNQVYESLPGFH- 1846

Db   2132 LALFELSCKLREYYDQLKVANFNRQECLCDKDEDMFVLRAGQGVVSGRN---SRLPLKHF 2188

Qy   1847 -----VYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDI 1901

Db   2189 KDHEFCFRSG----------GL----VPY---------------------DGTSRVDTI 2212

Qy   1902 LVFHDAINLIGALKVARCGMVGESFKSFEYK-------CYNAPPGGGKTTTLVDEFVK-- 1952

Db   2213 --FHTQTNFVSA-NALLSGYL--SYRTFTFTNLSANVLLYEAPPGGGKTTTLIKVFCETF 2267

Qy   1953 SPNSTATITANVGSSEDI-----NMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVL 2007

Db   2268 SKVNSLILTANKSSREEILAKVNRIVLDEGDTPLQTRDRILTIDSYLMN--NRGLTCKVL 2325

Qy   2008 -VDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAV------FVPKK 2060

Db   2326 YLDECFMVHAGAAVACIEFTKCDSAILFGDSRQIHYIDRNE---LDTAVLSDLNRFV-DD 2381

Qy   2061 ESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVV------SGKDKPVVRSLSKRPIG 2114

Db   2382 ESRVYGEVSYRCPWDVCAWLSTF--------YPKTVATTNLVSAGQS-----SMQVREIE 2428

Qy   2115 TTDDVAEINADVYLCMTQLEKSDM-----KRSLKGKGKETPVMTVHEAQGKTFSDVVLFR 2169

Db   2429 SVDDVEYSSEFVYLTMLQSEKKDLLKSFGKRSRSSVEKPT-VLTVHEAQGETYRKVNLVR 2487

Qy   2170 TKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDA 2215

Db   2488 TKFQEDDPFRSENHITVALSRHVESLTYSVLSSKRDDAIAQAIVKA 2533

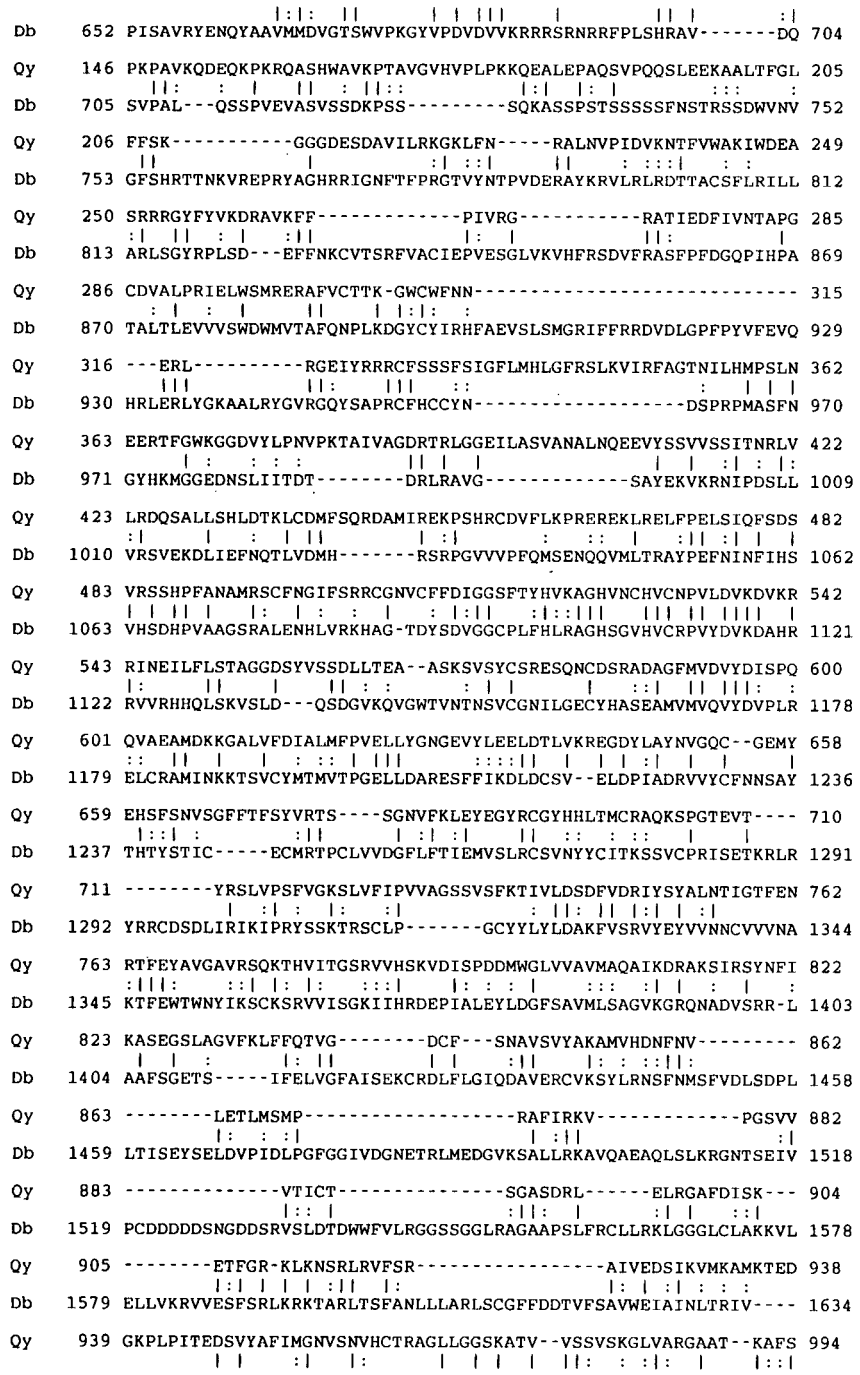

RESULT   6
Q66237
ID   Q66237      PRELIMINARY;      PRT;  3124 AA.
AC   Q66237;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   349-KDA VIRAL POLYPROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T36, QUICK DECLINE;
RX   MEDLINE=95266257; PubMed=7747424;
RA   Karasev A.V., Boyko V.P., Gowda S., Nikolaeva O.V., Hilf M.E.,
RA   Koonin E.V., Niblett C.L., Cline K., Gumpf D.J., Lee R.F.,
RA   Garnsey S.M., Lewandowski D.J., Dawson W.O.;
RT   "Complete sequence of the citrus tristeza virus RNA genome.";
RL   Virology 208:511-520(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T36, QUICK DECLINE;
RA   Karasev A.V.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U16304; AAC59623.1; -.
DR   MEROPS; C42.001; -.
DR   MEROPS; C42.002; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Polyprotein.
SQ   SEQUENCE   3124 AA;  349350 MW;  A02A517D7BDF6C7E CRC64;


  Query Match             6.4%;  Score 734;  DB 12;  Length 3124;
  Best Local Similarity   20.5%;  Pred. No. 1.9e-41;
  Matches  558;  Conservative  344;  Mismatches  959;  Indels  856;  Gaps  111;

Qy     47 SDFTPRCAAMIYIGKLTKGVKRTFVPPPVK------GFARQYAVVSGSVSALRGD----- 95

Db    594 SNFDELCHAV--LTRMSKGTTLEILLHSIRERKGLLGFAACFRWGSSHVFVKRGNFTARF 651

Qy     96 ----------GKKVLMEARTSTSATSDVSDFDVVFEAVSNALLVVHYHRVVPYAPVKREQ 145
```


Db 652 PISAVRYENQYAAVMMDVGTSWVPKGYVPDVDVVKRRRSRNRRFPLSHRAV-------DQ 704
Qy 146 PKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALTFGL 205
Db 705 SVPAL---QSSPVEVASVSSDKPSS---------SQKASSPSTSSSSSFNSTRSSDWVNV 752
Qy 206 FFSK----------GGGDESDAVILRKGKLFN-----RALNVPIDVKNTFVWAKIWDEA 249
Db 753 GFSHRTTNKVREPRYAGHRRIGNFTFPRGTVYNTPVDERAYKRVLRLRDTTACSFLRILL 812
Qy 250 SRRRGYFYVKDRAVKFF----------PIVRG----------RATIEDFIVNTAPG 285
Db 813 ARLSGYRPLSD---EFFNKCVTSRFVACIEPVESGLVKVHFRSDVFRASFPFDGQPIHPA 869
Qy 286 CDVALPRIELWSMRERAFVCTTK-GWCWFNN---------- 315
Db 870 TALTLEVVVSWDWMVTAFQNPLKDGYCYIRHFAEVSLSMGRIFFRRDVDLGPFPYVFEVQ 929
Qy 316 ---ERL----------RGEIYRRRCFSSSFSIGFLMHLGFRSLKVIRFAGTNILHMPSLN 362
Db 930 HRLERLYGKAALRYGVRGQYSAPRCFHCCYN----------DSPRPMASFN 970
Qy 363 EERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGEILASVANALNQEEVYSSVVSSITNRLV 422
Db 971 GYHKMGGEDNSLIITDT--------DRLRAVG----------SAYEKVKRNIPDSLL 1009
Qy 423 LRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVFLKPREREKLRELFPELSIQFSDS 482
Db 1010 VRSVEKDLIEFNQTLVDMH-------RSRPGVVVPFQMSENQQVMLTRAYPEFNINFIHS 1062
Qy 483 VRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKR 542
Db 1063 VHSDHPVAAGSRALENHLVRKHAG-TDYSDVGGCPLFHLRAGHSGVHVCRPVYDVKDAHR 1121
Qy 543 RINEILFLSTAGGDSYVSSDLLTEA--ASKSVSYCSRESQNCDSRADAGFMVDVYDISPQ 600
Db 1122 RVVRHHQLSKVSLD---QSDGVKQVGWTVNTNSVCGNILGECYHASEAMVMVQVYDVPLR 1178
Qy 601 QVAEAMDKKGALVFDIALMFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQC--GEMY 658
Db 1179 ELCRAMINKKTSVCYMTMVTPGELLDARESFFIKDLDCSV--ELDPIADRVVYCFNNSAY 1236
Qy 659 EHSFSNVSGFFTFSYVRTS----SGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVT---- 710
Db 1237 THTYSTIC-----ECMRTPCLVVDGFLFTIEMVSLRCSVNYYCITKSSVCPRISETKRLR 1291
Qy 711 --------YRSLVPSFVGKSLVFIPVVAGSSVSFKTIVLDSDFVDRIYSYALNTIGTFEN 762
Db 1292 YRRCDSDLIRIKIPRYSSKTRSCLP-------GCYYLYLDAKFVSRVYEYVVNNCVVVNA 1344
Qy 763 RTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRSYNFI 822
Db 1345 KTFEWTWNYIKSCKSRVVISGKIIHRDEPIALEYLDGFSAVMLSAGVKGRQNADVSRR-L 1403
Qy 823 KASEGSLAGVFKLFFQTVG--------DCF---SNAVSVYAKAMVHDNFNV--------- 862
Db 1404 AAFSGETS-----IFELVGFAISEKCRDLFLGIQDAVERCVKSYLRNSFNMSFVDLSDPL 1458
Qy 863 --------LETLMSMP----------RAFIRKV----------PGSVV 882
Db 1459 LTISEYSELDVPIDLPGFGGIVDGNETRLMEDGVKSALLRKAVQAEAQLSLKRGNTSEIV 1518
Qy 883 ----------VTICT----------SGASDRL------ELRGAFDISK--- 904
Db 1519 PCDDDDDSNGDDSRVSLDTDWWFVLRGGSSGGLRAGAAPSLFRCLLRKLGGGLCLAKKVL 1578
Qy 905 --------ETFGR-KLKNSRLRVFSR----------AIVEDSIKVMKAMKTED 938
Db 1579 ELLVKRVVESFSRLKRKTARLTSFANLLLARLSCGFFDDTVFSAVWEIAINLTRIV---- 1634
Qy 939 GKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATV--VSSVSKGLVARGAAT--KAFS 994
Db 1635 --------DGVVKTKLGCFRNLFRLIPKLYAGCKDVVKDVSNAYRSVVSFSAEVLGKSYS 1686
Qy 995 GITSFFSTGSLFYDRGLTEDERLD---------ALVRTENAINSPVG---------- 1032
Db 1687 KLCQFCLE---IFDIGLIEVSARNLGLETCVSIALAIFDVLITFPLGSVSFPVVCARVFG 1743
Qy 1033 --ILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKV----------L 1070
Db 1744 RLAMEVGSNFLFKKVFGKPETLGEDTFRRTVSMVLRPKAFDKLSVDAKGLVRCSGVIPGV 1803
Qy 1071 IGIFVASLGAAPIAW----------KYRRGIAANARRYAGSSYETLS 1107
Db 1804 LRIILSSLFSEDDAWVGYSKHEVSSLPIVTFCVKRLSYSLERCKEACVAFIVDSAKKILK 1863
Qy 1108 SLSSQ----------AAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTR-----ATVAKR 1150
Db 1864 SLTSEFNESIQNSDLTSGAKSLYGHVSSILYTSSGVINKAGSAVATVKAVGSDISGRVKR 1923
Qy 1151 QVPLALLS------FSTSYAISGCSML---------GIWAHALPRHLMFFFGLGTLLGAR 1195
Db 1924 MRPRFSLSRHFDEQTSEYYSASDCSELDESLFSETPGLRGNARCRDLF------RIVSRR 1977
Qy 1196 ASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPV 1255
Db 1978 AVA----------CC--------EYVTILLKRLASRAGEYSYSYVMDTLVRL-YV--- 2013
Qy 1256 EGRNVYDETLRYYRDFDYD----------EGAGPSGTQHEAVPGDDNDGST 1296
Db 2014 ----ATDDVVRRCRSFTFDCTDKASNFVTGFLSWRSLRSGA------YSATPG--LGGGS 2061
Qy 1297 SSVSSYDVVTN-VRDVGISTNG-EVTGEEE-----THSPRS------VQ---------- 1332
Db 2062 RGLTSFSHTGGFTRVIGALGNTLSVVFDEEFFRTLLHRLRADVVVLLVEFLKNLPVCLFS 2121
Qy 1333 -YTYVE----------EEVAPSAAVAERQGDPSGSGTADAMAFVESVKKGV 1372
Db 2122 IYTFIECRGRIFPLREFLWGFCREVWDVIEGLSINTYQ---------YLSAVAEKVYRGA 2172
Qy 1373 DDVFHQQSSGETAREV-EVDGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQ 1431
Db 2173 FNRRFNEEDERTLRVITEAEGRIEFLRVVLAE---MERFRAASNEVS---------- 2216
Qy 1432 SSLVSSPQADIPKV--TQSEVHAQKEVKQEVPLA----TVSGATPIVDEKPAPSVTTRGV 1485
Db 2217 ----SSRSSDSPSFPPTYGEIEELNDDSSDSALVIGVDNGTGRTIVDGETDSASGSDSDV 2272
Qy 1486 KI-IDKG----------KAVAHVAEKKQVQVEQPKQRSLTINEGKA 1520
Db 2273 SIPLDRDFEQIGRLGGNGNVDVVRLGLKFVLKLLRNLFDARSFKVLYGLLLQTLFYGGSV 2332
Qy 1521 GKQLC----MFRTC--------SCGV----------QLDVYNEATIATRFSNAFTFVD 1556
Db 2333 NKFVCRLVFSLGTCLPFELLRFACSARRLLKSLASFVYQTNLFSEVMVEHLFSVSRLLYD 2392
Qy 1557 NLKGRSAVF----------FSKLGEGYTY-NGGSHVSSGWP----------RAL 1589
Db 2393 S-RYFATVFSELERLNFYLDYFLSLPKRYLFPNSRARSCARVPVPVIRTEGVAPVRERSL 2451
Qy 1590 EDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRK 1649
Db 2452 NDILDSLDSLKVSD--------LKGKSTFVDDGESDSGSEHSGSSISF-GSVNEE----- 2497
Qy 1650 GGKVMVINVASGDYFLMPCGFQRTHLHSVNSID---EGRISLTFRATRRVFGVGRMLQLA 1706
Db 2498 -----VLNAVNGDSVCEKAPLMK-HVTALSRLERHPSGRKDIP--VSRKTFVRGESSGVK 2549
Qy 1707 GGVS--------DEKSPGVPNQQPQSQGATRTITPKSGGKALSEG------SGREVKGRS 1752
Db 2550 SKVSMCDYLVTLNSSFGTVPDMYP----VARDITYKKLTNAMREFYYSQKITLYELHGKL 2605
Qy 1753 TYSIWCE---QDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIG 1809
Db 2606 S-SYWDELKVAGFDRKLTRMDLDDDVY----------VVDF------- 2635

```
Qy   1810 IERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHVYKSG-TDLIFHSTQDGL----R 1864
Db   2636 ----NRKLLVGRN--------------------GSHPFRNTLKSYCFMFCNDGLVPCPR 2670

Qy   1865 VRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCD--DILVFHDAINLIGALKVARCGMV 1922
Db   2671 SQKYDYALISNQTSF---------VAANSFLRAVDGKDLTFTNEEHSLI---------- 2710

Qy   1923 GESFKSFEYKCYNAPPGGGKTTTLVDEF----VKSPNSTATITANVGSSEDINMAVKKRD 1978
Db   2711 ----------VYEAPPGGGKTHSLVNSYADYCVKV--SCLVVTANKNSQTEISQRISN-- 2756

Qy   1979 PNLEGLNSAT----------TVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATG 2027
Db   2757 -ELMGRKLAAKYVTDAASRVFTVDSYLMNHL-RLTTQLLFIDECFMVHAGAIGAVVEFTS 2814

Qy   2028 ASEGLFFGDINQIPFINREKV---FRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMT 2084
Db   2815 CKAVVFFGDSKQIHYIHRNDLGVSFVADIDAFI-QPEHRIYGEVSYRCPWDICEWLSEF- 2872

Qy   2085 VRGTEKCYPEKVVSGKDKPVVR-SLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSL- 2142
Db   2873 -------YPRHVATANVGSIGKSSVSIEEINGCDDVPYDKAAKYIVYTQAEKNDLQKHLG 2925

Qy   2143 ---KGKGKETPVM-TVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYA 2198
Db   2926 RLTVGRNKVVPIVNTVHEVQGETYKRVRLVRFKYQEDTPFSSKNHIVVALTRHVDSLVYS 2985

Qy   2199 ALSSKLDDKVGTYISDA 2215
Db   2986 VLTSRRYDDTATNIDRA 3002
```

RESULT 7
O40959
ID O40959 PRELIMINARY; PRT; 2301 AA.
AC O40959;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE METHYLTRANSFERASE (MT) AND HELICASE (HEL) DOMAINS.
OS Little cherry closterovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=53180;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UW2;
RX MEDLINE=97410293; PubMed=9267008;
RA Jelkmann W., Fechtner B., Agranovsky A.A.;
RT "Complete genome structure and phylogenetic analysis of little cherry
RT virus, a mealybug-transmissible closterovirus.";
RL J. Gen. Virol. 78:2067-2071(1997).
DR EMBL; Y10237; CAA71285.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Transferase; Methyltransferase; Helicase.
SQ SEQUENCE 2301 AA; 259625 MW; 2343F291C34E4E45 CRC64;

Query Match 6.3%; Score 721; DB 12; Length 2301;
Best Local Similarity 19.3%; Pred. No. 8.7e-41;
Matches 470; Conservative 350; Mismatches 874; Indels 742; Gaps 93;

```
Qy     68 RTFVPPPVKGFARQYAVVSGSVSALRGDGKKVLM--EARTSTSATSDVSDFDVVFE---A 122
Db    337 RTLRRPASTGYVR---------------KPVLRAHQAAQTAKIFSSFSEFSIALEKQKA 380

Qy    123 VSNALLVVHYHRVVPYAPVKREQPKPAVKQDEQKPKRQASHWAVKPTAVGVHVPLPKKQE 182
Db    381 ASTASAI----KFGDFEVFDLREPDNVIKSYAQ---ALTSNLPVDDRILKLNIPVSSEVK 433

Qy    183 ALEPAQSVPQQSLEEKAALTFGLFFSKGGGDESDAVILRKGKLFNRALNVP--------- 233
Db    434 TECPSTMNPSNPKSEIAKRAVEL-SSRATSDSSGEIY---SKTFNQLLNIPTGGTIIRHF 489

Qy    234 -----------------IDVKNTFVWAKIWDEASRRRGYFYVKDRAVKFFPIVRGRATIE 276
Db    490 INGVCRSFTLNRSDGRVFNILNNRNAAKIMFNLSLKHKKFYV------------------ 531

Qy    277 DFIVNTAPGCDVALPRIELWSMRERAFVCTTKGWCWFNNERLRGEIYRRRC-FSSSFSIG 335
Db    532 DGYCRTPSG---------------KLFSEYDNGFCWLQAFAMFGKIIPTFVEFIPNLNVS 576

Qy    336 FLMHLGF--RSLKVIRFAGTNILHM-PSLNEERTFGWKGGDVYLPNVPKTAIVAGDRTRL 392
Db    577 CLLQAGLPKSCLKYMHKTGNALYHFDPN---------KINNIYHANL---NFLVGSAMNE 624

Qy    393 GGEILAS--------------VANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLC 438
Db    625 NKDILESDNIDLIVEGFTKQVIAKAYPKSD--NQAVTNILNRVGCR--------INT-WC 673

Qy    439 DMFSQRDAMIREKPSHRCDVFLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFN 498
Db    674 D--RKNDLMI--------FVHLSTRQKKIVCDLFPHLKFEFKETTQSSHPVFNVVRTVSN 723

Qy    499 GIFSRRCGNVCFFDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDSY 558
Db    724 FVLYKMQEGRHFVDFGGNIGTVINSECDDVHICNPVADSRDAKRHVDNGLFLAKSVG--- 780

Qy    559 VSSDLLTEAASKSVSYCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIAL 618
Db    781 ---------VSNNISVCNKLAQHCNHKSDRAVMVEVYDMTLTEMCQAMLAHGTIRLDFIL 831

Qy    619 MFPVELLYGNGEVYLEELDTLVKREGDYLAYNVGQCGEMYEHSFSNVSGFFTFSYVRTSS 678
Db    832 LLPGDLLEDFNTITIFDGGCKITKDDDKVYYYYGDAAEAYTHDLNNLRNIMTDNLV-CVD 890

Qy    679 GNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSF-VGKSLVFIPV-VAGSSVSF 736
Db    891 GTAFKKTLETSYGPFRHFSLTKLETFPSGKIEFLTMYDKCEKNKMLVKVPMRNINGMVTL 950

Qy    737 KTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDD 796
Db    951 HDLIVDRSMLMNLIEYCANCVENFNRKGFEHIMSQYRSRKNYVIYNGKVIHDQVDIPLDL 1010

Qy    797 MWGLVVAVMAQAIK--DRAKSIRSYNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKA 854
Db   1011 LPGFIGVCLSEGIRLAEKVHYLAKYSYYQ--------------------------YYAPG 1044

Qy    855 MVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNS 914
Db   1045 I----FNCM--IYSIRQLFLR------FKTVCYEKVLDLMDF-----LFGEWITTTARNS 1087

Qy    915 RLRVF---SRAIVEDSIKVMKAMKTEDGKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGS 971
Db   1088 GTKIFDISASVVVKQSLLI-----SADG------------------DVH---------- 1113

Qy    972 KATVVSSVSKGLVARGAATKAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINSPV 1031
Db   1114 --------------------------------SIF---------------TEN------ 1119

Qy   1032 GILETSRVAVSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGI 1091
Db   1120 ----------------------------FEKYQQR----------------------- 1126

Qy   1092 AANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQ 1151
Db   1127 CADLEQHA-SEYDTASELSDAIDSGGMGVSDLKI----------FEKFIT---------- 1165

Qy   1152 VPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASANTWKFGGF----S 1207
Db   1166 -KLAELGFG-----------------ERNVL----------RFSAKLYSCFDWLRRQS 1195
```

```
Qy   1208 NNWCAVPEVVWRGKSVSSL--LLPITLGVSLIIR---------------GLLN------D 1244
          ||   :  |:   ||: ||  |  | :   |:|:              ||||      |
Db   1196 NNDSVLRRVIGYFKSLFSLSTLKNILISPYLLIKWICDCGKASFKTFVSGLLNVFQTFSD 1255

Qy   1245 TIPQLAYVPPVEGRNVYDETLRYYRDFD---YDEGAGPSGTQHEAVPGDDNDGSTSSVS- 1300
           |  |  :       :  :| :  :|: |   :|:  |   |::  |   |:      :|
Db   1256 RINFLVKIALDSDEDFQNEVVAAFRNHDLEAFDDTVGKVFTKNFDVVVSDDFTKKKIISI 1315

Qy   1301 -----SYD--VVTNVRDV----------------GISTNGEVTGEEETHSPRSVQYTYV 1336
               ||:  :  :|:||                  ::  | :    |      |    :
Db   1316 IEDELSYERGIDVDVQDVKKEILEIASSADTPTITKVNNLGHLDDALENDVRHLVLDIQL 1375

Qy   1337 EEEVAPSAAVAERQGDPSGSGTADAMAFVESV----------------KKGVD--DVFH 1377
           :||     | |   :   || |      : :                 | | |  |: |
Db   1376 RKEVL--RMVDEDLNEALESGGAGKSKLCDKIFNYILDRLNLFKHYITSKFGWDFYDILH 1433

Qy   1378 QQSSGETAREVEVDGKGLL--PESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLV 1435
               |:  ||:|    :|      : :  |       |       :|:  || : : |:
Db   1434 -----ESFEEVKVRKSAILCAYHDDLSDIYT------DNGLVLKPLNKSAREVLYARLI 1481

Qy   1436 SSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVT-TRGVKIIDKGKAV 1494
          :                | : :      :             ||:    |:    | | :
Db   1482 AK--------------MKFIYRNFMTTFI-----------PSMRFLSGI----KNKIM 1510

Qy   1495 AHV--AEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAF 1552
          :|:  |||:|   ::    :||:             :|:   ||           |   |
Db   1511 SHICDAEKRQAIFDKCLDATLTV------------SCTATAQL-----------FGGNF 1546

Qy   1553 TFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRALEDILTAIKYPSVF--DHCLVQKY 1610
            :  :   | :  : | : :||  | |          : |:: |   ::    :  | |
Db   1547 CLIRLI--LSPILSAYLKKSWTYYFGEH---------DGIISTIIASTLLGGGNYSVNKM 1595

Qy   1611 KMGGGVPFHADDEEC-----YPSDNPILTVNLVGKANFSTKCRK---------------- 1649
           :  |  ||     |     | |    :| : : : || |  :
Db   1596 AIVSGSYFHFKPYFCNKAMKYNSFKQYVT-DYIARHNFFTTAKTTVGLITSPVGVAVLTV 1654

Qy   1650 -------------------------GGKVMVIN--VASGD--YFLMPCGFQRTHLHSVNS 1680
                                    | | | |  :|:||    | | |  :  :: |
Db   1655 GALSGINPFAAMFFFAAASHYKNFYDGIVRVSNIAIATGDTLLHLKPVGALKKAINKVKR 1714

Qy   1681 IDEGRISLTFRATRRVFGVGRMLQLAGGVSDEKSPGVPN---------------------- 1719
          :     ::|  :| || |     :|    :| |:    |
Db   1715 MKFKNSNITSNST-RVNEVNNDGHIAND-TDNKNSETSNIAYEEVEEQSDDDEIVDIYMK 1772

Qy   1720 ------------QQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKC 1766
                       | |  |  : :|       | : |  :    |:||
Db   1773 RVKKDKMPIKEYIDQSQEPGVKQALT-----NQLLKDSYNDPNASSSYSTGA-------- 1819

Qy   1767 EWLRADNPVMALEPDYTPM------TFEVVKTGTSEDAVVEYLKYLAI-GIERTYRALLM 1819
                       |: |:       |:: |     |  |:  : | :  | | | |
Db   1820 ------------VDHLPVDGKYQSVFDIRKNHRGPDISVQLCEELPMCEIMRNYPA--- 1863

Qy   1820 ARNIAVTTAEGVLKVPNQVYESL----------------------PGFHVYKSGTDLIFH 1857
          ::||     : | :| | : | :                     |   |:    |:
Db   1864 SKNIVF--QQTVNEVENSIQEFVQIESLTLLLNFKKLETTLDKFDSGMKTYRQLQTLVND 1921

Qy   1858 STQD-----------GLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNL-FVCDDILVFH 1905
          |||              |   |:|  :  |  : |  : :    : || : |   :
Db   1922 STQFMYVKGQDWVPLSLIHSDVPQTV--EYSV-TSSRKLARFDSNGDEVQFTTTEFAAGF 1978

Qy   1906 DAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTT----TLVDEFVKSPNSTATIT 1961
               ||    |  :   : :| :    |  | ||| ||||     : |  ::      |  :
Db   1979 SNKKLISMDKFRQSANIIDSDRLSTMKIVNKPPGSGKTTQISQMMFDCHIRGKTCLALTS 2038

Qy   1962 ANVGSSEDI----NMAVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQG 2017
            || :| |    |: |  :  |:: :::    |      ||  :|    ||| :| | |
Db   2039 TKVGKNELIEKLKNLIVGAKPKNVKTVDA---FNMHGKKAIVDELY----VDECYMSHSG 2091

Qy   2018 LLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKESVVYTSKSYRCPLDVC 2077
          : : :     :   |||| ||||: |   |    |  | |   | : | ||||| |||
Db   2092 AILMALSNVDFTNCKFFGDENQIPYSVRLTGFEALFATTVFKYCHVTHDSISYRCPADVC 2151

Qy   2078 YLLSSMTVRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSD 2137
          ||||  | |   | |   | :  ::  :|::    :    ::      ||   || || :
Db   2152 YLLSHATDRAGNKMYTNGVYAVGNQN-IRTMHLNALTGVANIPIREDTVYAAFTQGEKDE 2210

Qy   2138 MKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVY 2197
          : |      |:    ||:| || || ::::|| :  :: |:      :  :|||| :  |
Db   2211 LNRF----AKKNIAKTVNELQGGTFKNLIVFRLRLHNNVLYNDTRQFITAISRHTETFNY 2266

Qy   2198 AALSSKLDDKVGTYISDASPQSVSDALLHTFAPAGC 2233
               |: ::| :   |:  :  :: | ::  |    |
Db   2267 FCPSNAMNDSISNAINKLN--TIEDNVVAQFGFRQC 2300


RESULT   8
Q08534
ID   Q08534      PRELIMINARY;      PRT;  3074 AA.
AC   Q08534;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (CONTAINS: 66 KDA
DE   PROTEIN).
OS   Sugar beet yellow virus (SBYV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=31714;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UKRAINIAN (BYV-U);
RX   MEDLINE=94082464; PubMed=8259666;
RA   Agranovsky A.A., Koonin E.V., Boyko V.P., Maiss E., Froetschl R.,
RA   Lunina N.A., Atabekov J.G.;
RT   "Beet yellows closterovirus: complete genome structure and
RT   identification of a leader papain-like thiol protease.";
RL   Virology 198:311-324(1994).
CC   -!- FUNCTION: PROBABLE POLYMERASE.
CC   -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY A RIBOSOMAL FRAMESHIFT
CC       THAT OCCURS BETWEEN POSITIONS 2630 AND 2631.
DR   EMBL; X73476; CAA51871.1; -.
DR   MEROPS; C42.001; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR001788; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF00978; RNA_dep_RNApol2; 1.
DR   PFAM; PF01443; Viral_helicase1; 2.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   RNA-directed RNA polymerase.
FT   CHAIN         1   3074       348 KDA PROTEIN.
FT   CHAIN         1    588       66 KDA PROTEIN.
FT   MUTAGEN     509    509       C->T: NOT PROCESSED INTO ITS SUBUNITS.
FT   MUTAGEN     517    517       C->T: REDUCED PROCESSING INTO SUBUNITS.
FT   MUTAGEN     518    518       C->T: REDUCED PROCESSING INTO SUBUNITS.
FT   MUTAGEN     556    556       H->E: NO EFFECT.
FT   MUTAGEN     569    569       H->E: NOT PROCESSED INTO ITS SUBUNITS.
FT   MUTAGEN     588    588       G->D: NOT PROCESSED INTO ITS SUBUNITS.
SQ   SEQUENCE   3074 AA;  346028 MW;  D75B9564658BEB0C CRC64;

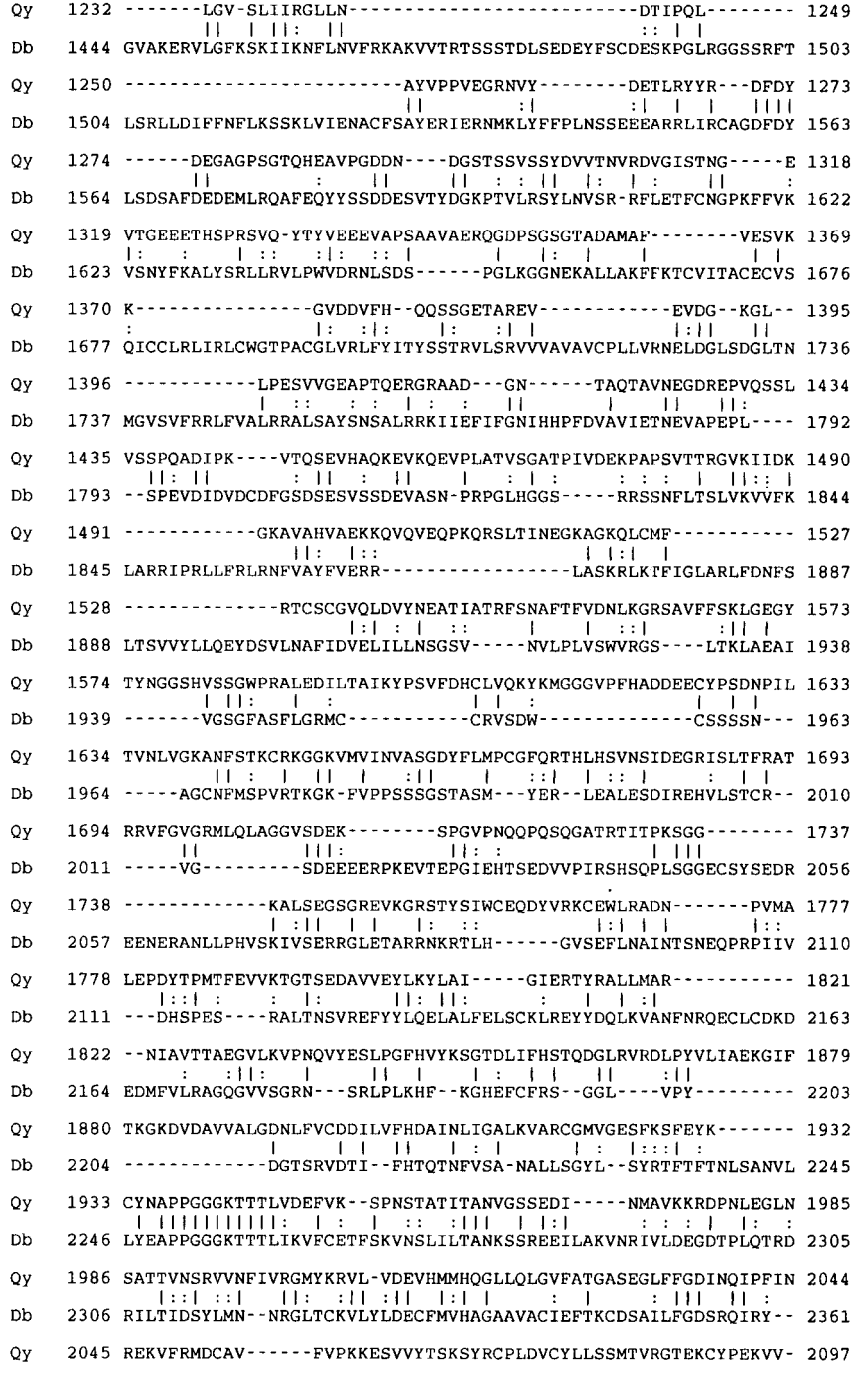

  Query Match             6.3%;  Score 721;  DB 12;  Length 3074;
  Best Local Similarity   21.8%;  Pred. No. 1.6e-40;
  Matches  550;  Conservative  318;  Mismatches  898;  Indels  762;  Gaps  117;

Qy     137 PYAPVKREQPKPAVKQDE-----QKPKRQASHWAVKPTA-VGVHVPLPKKQEALEPAQSV 190
           |  |  ||   | ||:        || |       :| |  |       : |  :  |
Db     319 PRLPKFREVEIPVVKKHAVPAVVSKPVR-----TFRPVATTGAEYVNARNQCSRRPR--- 370

Qy     191 PQQSLEEKAALTFGL------FFSKGGGDESDAVILRKGKLFN--RALNVPID-----VK 237
                 :   |: |||        | |    |      ::: |: :       |:     :|
Db     371 -NHPILRSASYTFGFKKMPLQRFMK----EKKEYYVKRSKVVSSCSVTKSPLEALASILK 425

Qy     238 NTFVWA------KIWDEASRRRGYFYVKDRAVKFFPIVRGRATI-------EDFIVNTA- 283
           |   ::      | :|       :|   |  ::  |:  |: ::       | : | ||
Db     426 NLPQYSYNSERLKFYD-------HFIGDDFEIEVHPLRGGKLSVLLILPKGEAYCVVTAA 478
```





```
Db   2362 -GRCSELDTAVLSDLNRFV-DDESRVYGEVSYRCPWDVCAWLSTF--------YPKTVAT 2411

Qy   2098 -----SGKDKPVVRSLSKRPIGTTDDVAEINADVYLCMTQLEKSDM-----KRSLKGKGK 2147

Db   2412 TNLVSAGQS-----SMQVREIESVDDVEYSSEFVYLTMLQSEKKDLLKSFGKRSRSSVEK 2466

Qy   2148 ETPVMTVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDK 2207

Db   2467 PT-VLTVHEAQGETYRKVNLVRTKFQEDDPFRSENHITVALSRHVESLTYSVLSSKRDDA 2525

Qy   2208 VGTYISDA 2215

Db   2526 IAQAIVKA 2533


RESULT   9
Q9WID7
ID   Q9WID7      PRELIMINARY;      PRT;  3115 AA.
AC   Q9WID7;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P349 PROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99190444; PubMed=10092023;
RA   Vives M.C., Rubio L., Lopez C., Navas-Castillo J., Albiach-Marti M.R.,
RA   Dawson W.O., Guerri J., Flores R., Moreno P.;
RT   "The complete genome sequence of the major component of a mild citrus
RT   tristeza virus isolate.";
RL   J. Gen. Virol. 80:811-816(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Moreno P.;
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y18420; CAA77161.1; -.
DR   INTERPRO; IPR000515; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
SQ   SEQUENCE   3115 AA;  346982 MW;  3EEA612F605D5045 CRC64;


  Query Match             6.1%;  Score 706.5;  DB 12;  Length 3115;
  Best Local Similarity   20.1%;  Pred. No. 1.7e-39;
  Matches  482;  Conservative  314;  Mismatches  826;  Indels  777;  Gaps   97;

Qy    298 MRERAFVCTTKGWCWFNNERLRG------------EIYRRRCFSSS----FSIGFLMHL 340

Db    891 IRHIAEVCLSMGKVFFRRDVNLGAFPYVFEVQHRLERLYGKRAVEYSVRGYFSAPRLYHC 950

Qy    341 GFR-SLKVIRFAGTNILHMPSLNEERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGEILAS 399

Db    951 CFTDSPRPFRF----------LNKYHKMGGEEGALLITDT--------DRLRAVGSVYDK 992

Qy    400 VANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVF 459

Db    993 VKKALPD----SLLVRSVEKDLIEFNQTL----VDTQ-----------RARPGVAVPFQ 1032

Qy    460 LKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTY 519

Db   1033 MSENQQVMLTRAYPEFNINFTHSVHSDHPVAAGSRMLENHLVRKYAGS-DYSDVGGCPLF 1091

Qy    520 HVKAGHVNCHVCNPVLDVKDVKRRI------NEILFLSTAGGDSYVSSDLLTEAASKSVS 573

Db   1092 HLRAGHTGVHVCRPVYDVKDAHRRVVRHHQLGKVAFDHTDGSKQV--------GAVNTNT 1143

Qy    574 YCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYL 633

Db   1144 VCGNVLGECFHASEAIVMVQVYDVPLRELCRAMVNKKTSVCYMTMVTPGELLDARESFFI 1203

Qy    634 EELDTLVKREGDYLAYNVGQC--GEMYEHSFSNVSGFFTFSYVRTS----SGNVFKLEYE 687

Db   1204 KELDCSV--ELDVVADRVVYCFNNSAYTHAYSTIR-----EYMRTPCVVVDGFLFTIEMV 1256

Qy    688 GYRCGYHHLTMCRAQKSPG-TEV-----------TYRSLVPSFVGKSLVFIPVVAGSSVS 735

Db   1257 SLRCSVNYYCITKSRVCPRISEVKRLRYRRCDSDLIRIKIPRYSSKTRSCLP-------G 1309

Qy    736 FKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPD 795

Db   1310 CYYLYLDAKFVSRVYEYVVNNCVVVNAKTFEWTWNYIKSCKSRVVISGKVIHRDVPIALE 1369

Qy    796 DMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLF-FQTVGDC------FSNAV 848

Db   1370 YLDGFSAVMLSAGVKGRQNAEAFSRRLAAFSGD-TSLFELVRFAVSEKCRDLFVGIQEAV 1428

Qy    849 SVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASD-RLELRGAFDISKETF 907

Db   1429 ERCVKSYLRNSFNMSFVDLSDP----------LLTISEYSELDVPIDLPGFGCITEEDE 1477

Qy    908 GRKLKNSRLRVFSRAIVEDSIK---VMKAMKTE-----------------DGKPLPIT 945

Db   1478 AK-------------LIEDGVKSALLRKAVRDEAQLILERGGTSEIVPCGDDDGLGSDSS 1524

Qy    946 EDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGL-----VAR----------GAA 989

Db   1525 DSSLGS---GSDSSSKASSGGLRAGAAPSLFRSLLKKLGGGFCVARRLLDLLCERLVGSL 1581

Qy    990 TKAFSG---ITSPFS-----TGSLFYDRGLTED--ERLDALVRTENAINSPVGILETSRV 1039

Db   1582 SRLKFGTTHLTSFFSMLRAKLSCNFFDDSIFTAVWDTVMNLIRIVNGVTKTK--LECFRN 1639

Qy   1040 A---VSKVVAGTKEFWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANAR 1096

Db   1640 VLRLIPKLCSGCKDF-----AKDFADMC----------------------KSVLLFST 1670

Qy   1097 RYAGSSYETLSSLSSQAAG-GLRGLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLA 1155

Db   1671 EFFSKSYTKLCRFSLEIFDVGLIEISARNM---------GLETCVSIASAIF---DVILT 1718

Qy   1156 LLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLG--TLLGARASANTWKFGGFSNNWCAV 1213

Db   1719 YPLGSISFPVVCARVFGALAMEVGSNYCFKKTLGKPEILGEDAFRRT------------- 1765

Qy   1214 PEVVWRGKSVSSLLLPITLGVSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDY 1273

Db   1766 VSMVLSAKAFDKL--------SVDAKGLV-----RCSGIIPGVLRIVFSSLFS-----DA 1807

Qy   1274 DEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQY 1333

Db   1808 DAWVGYS--KHE------------VSSLPTI------------------TYSAKRLSF 1833

Qy   1334 TYVEEEVAPSAAVAER----------------QGDPSGSGTADAMAFVESVKKGVDDVFH 1377

Db   1834 VLEQTKEACIAFIVDSSKRILASLTSEIHESLQNSELTSGAKSLYGRVSSILSASNGMVN 1893

Qy   1378 QQSSGETAREVEVDGKGLLPESVVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSS 1437

Db   1894 K--AGNAVAAVKAAGSG-LSKRVKRIRPSFHLARCVNEQTSEYYSASDDSELDESLLSDT 1950

Qy   1438 PQA-------DIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDEKPAPSVTTRGVKIIDK 1490

Db   1951 PGLRGNGKCNDLFRVVSGRVLVYRR-------------------------RAVEALK- 1982

Qy   1491 GKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFRTCSCGVQLD--VYNEATIATRF 1548

Db   1983 -RLILHVGECLYSRIADSFVRLFVRTNTAVGK--CESLAVDCATKITDLVVGYITWRNLH 2039
```

```
Qy   1549 SNAFTFVDNLKGRSAVFFSK-LGEGYT------------YNGGSHVSSGWPRALEDILT 1594
Db   2040 FGAYSTTPGLSGGSRVSLSRLLGDGLTCVIQNFVTAWLTIFNETFYLSLAHRMRTASINV 2099

Qy   1595 AIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRK----- 1649
Db   2100 LLEFLKNLPVCLVSIYTF------------------------------IECRSRIFPL 2127

Qy   1650 -----GGKVMVINVASGDYFLMPCGFQ----------RTHLHSVNSIDEGRISLTFRAT 1693
Db   2128 KEFLWGFCIFVWNIIKE---LSVCIYENLPVCITRLCNRAFRRSSNEDDERTLQVITEAE 2184

Qy   1694 RRVFGVGRMLQLAGGVSDEKS------PGVPNQQ--PQSQG----------ATRTITPK 1734
Db   2185 NR---IGFLREVLAEMARYRASLEIPVPGSPNSPLFPPTYGEIEELGDGSSDSARVIGVD 2241

Qy   1735 SGGKALSE----------------------------GSG-------------------- 1745
Db   2242 NGSDSMSTEGEAELDSDGDSDDSIPLDRDFEQIGGLGGSGNVDVFRLGLRFVWRMSRNLF 2301

Qy   1746 --REVK------------GRST--------YSI----------------W---------- 1757
Db   2302 NLRAVKFSYGLALQCIFNGRSTGQFISRLLFSFGGCLPFEFFRHVCAVRWLSRSLALHVR 2361

Qy   1758 ----CEQDYVRKCEWLRADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERT 1813
Db   2362 NASICDEETV---EWLFSFSRFL-YDSRYLSLVFSEL------ERVEFYLSYLFSLPKRC 2411

Qy   1814 Y----------RA-LLMARNIAVTTAEGVLKVPNQVYESLPGFHV-YKSGTDLIFHS--- 1858
Db   2412 FTHNRSVQVCVRAPVRVVRNVGL--VEDGRKSLSEVLNSLDSLKVSNRKGKSAVIESDSS 2469

Qy   1859 ----------------------TQDGLRV----------------------RDLPYVLI 1873
Db   2470 DSESGSSSGSISFGSINEEVLKTANGSVVSEKASLMKHAAALSGLEPDLGGKKDLPII-- 2527

Qy   1874 AEKGIFTKGK---------------------------------------------------- 1883
Db   2528 --RRNFTRGESSGTKSVGSLCEYLTTLNSSFGSVPDLYPVARDITYKKMTNAMREFYYSQ 2585

Qy   1884 ------------------------------DVDAVVALGD-NLFV--------------- 1897
Db   2586 KVTLYELHGKLSSYWDELKTVGFDRKLAKMDLDDDVGVVDFNLKVIVGRHGTRPFRNTVN 2645

Qy   1898 ------CDDILV------------FHDAINLIGALKVARCGMVGE--SFKSFEYK--CYN 1935
Db   2646 SHEFMFCDDGLVLCPGPKKRDFALISSQTQFVAANSFLR-AVDGKDLTFTNSEHSLIVYE 2704

Qy   1936 APPGGGKTTTLVDEFVK--SPNSTATITANVGSSEDINMAVKKRDPNLEGL--------N 1985
Db   2705 APPGGGKTHSLVTYFADYCTKVSCLVVTANKNSQAEISQRISREFMDRKMLAKHVVKTAG 2764

Qy   1986 SATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINR 2045
Db   2765 RVFTVDSYLMNHL-RLTTDLLFVDECFMVHAGAIGAVVEFTSCKAVVFFGDSRQIHYIHR 2823

Qy   2046 EKV---FRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDK 2102
Db   2824 NDLGVSLLHDIDAFI-EPQHRIYGEVSYRCPWDICEWLS--------KFYPRHVATTNTG 2874

Qy   2103 PVVR-SLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSL----KGKGKETPVM-TVHE 2156
Db   2875 SVGKSSVSIEVINGCDDVPYDSSAKYIVYTQAEKNELQKHLGRLTVGRTKAVPVVNTVHE 2934

Qy   2157 AQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDA 2215
Db   2935 VQGETYKKVRLVRCKYQEDTPFCSDNHVVVALTRHVDSLTYSVLNSRRYDKTASSIDEA 2993


RESULT 10
Q9IFX0
ID   Q9IFX0      PRELIMINARY;      PRT;  3115 AA.
AC   Q9IFX0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   P349 PROTEIN.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T30;
RA   Albiach-Marti M.R., Mawassi M., Gowda S., Satynarayana T., Hilf M.E.,
RA   Shanker S., Almira E.C., Vives M.C., Lopez C., Guerri J., Flores R.,
RA   Moreno P., Garnsey S.M., Dawson W.O.;
RT   "Sequences of Citrus tristeza virus separated in time and space are
RT   essentially identical.";
RL   J. Virol. 74:6856-6865(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T30;
RA   Albiach-Marti M.R., Dawson W.O.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF260651; AAF70348.1; -.
SQ   SEQUENCE   3115 AA;  346874 MW;  6B97F494ED25AD63 CRC64;


  Query Match             6.0%;  Score 689.5;  DB 12;  Length 3115;
  Best Local Similarity   20.8%;  Pred. No. 2.7e-38;
  Matches  487;  Conservative  316;  Mismatches  880;  Indels  655;  Gaps   97;

Qy    298 MRERAFVCTTKGWCWFNNERLRG------------EIYRRRCFSSS----FSIGFLMHL 340
Db    891 IRHIAEVCLSMGKVFFRRDVNLGAFPYVFEVQHRLERLYGKRAVEYSVRGYFSAPRLYHC 950

Qy    341 GFR-SLKVIRFAGTNILHMPSLNEERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGEILAS 399
Db    951 CFTDSPRPFRF----------LNKYHKMGGEEGALLITDT--------DRLRAVGSVYDK 992

Qy    400 VANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHRCDVF 459
Db    993 VKKALPD----SLLVRSVEKDLIEFNQTL----VDTQ------------RARPGVAVPFQ 1032

Qy    460 LKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGGSFTY 519
Db   1033 MSENQQVMLTRAYPEFNINFTHSVHSDHPVAAGSRMLENHLVRKYAGS-DYSDVGGCPLF 1091

Qy    520 HVKAGHVNCHVCNPVLDVKDVKRRI------NEILFLSTAGGDSYVSSDLLTEAASKSVS 573
Db   1092 HLRAGHTGIHVCRPVYDVKDAHRRVVRHHQLGKVAFDHTDGSKQV--------GAVNTNT 1143

Qy    574 YCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGEVYL 633
Db   1144 VCGNILGECFHASEAIVMVQVYDVPLRELCRAMVNKKTSVCYMTMVTPGELLDARESFFI 1203

Qy    634 EELDTLVKREGDYLAYNVGQC--GEMYEHSFSNVSGFFTFSYVRTS----SGNVFKLEYE 687
Db   1204 KELDCSV--ELDVVADRVVYCFNNSAYTHAYSTIR-----EYMRTPCVVVDGFLFTIEMV 1256

Qy    688 GYRCGYHHLTMCRAQKSPG-TEV-----------TYRSLVPSFVGKSLVFIPVVAGSSVS 735
Db   1257 SLRCSVNYYCITKSRVCPRISEVKRLRYRRCDSDLIRIKIPRYSSKTRSCLP-------G 1309

Qy    736 FKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPD 795
Db   1310 CYYLYLDAKFVSRVYEYVVNNCVVVNAKTFEWTWNYIKSCKSRVVISGKVIHRDVPIALE 1369

Qy    796 DMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKLF-FQTVGDC------FSNAV 848
Db   1370 YLDGFSAVMLSAGVKGRQNAEAFSRRLAAFSGD-TSLFELVRFAVSEKCRDLFVGIQEAV 1428

Qy    849 SVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFG 908
Db   1429 ERCVRSYLRNSFNMSFVDLSDP-----------LLTI-----SEYSELDVPIDL--PGFG 1470
```

```
Qy    909 RKLKNSRLRVFSRAIVEDSIK---VMKAMKTE---------DGKPLPITED-------SV 949
Db   1471 CITEGDEAK-----LIEDGVKSALLRKAVRDEAQLILERGGTSEIVPCGDDDGLGSDGSD 1525

Qy    950 YAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGL-----VAR-----------GAATKAF 993
Db   1526 SSLGSGSDSSSKASSGGLRAGAAPSLFRSLLKKLGGGFCVARRLLDLLCERLVGSLSRLK 1585

Qy    994 SG---ITSFFS----------TGSLF-------------------------------- 1006
Db   1586 FGTTHLTSFFSMLRAKLSCNFFDDSIFTAVWDTVMNLIRIVNGVTKTKLECFRNVLRLIP 1645

Qy   1007 ----------------------------------------YDRGLTE----DERLDAL 1020
Db   1646 KLCSGCKDFAKDFADMCKSVLLFSTEFFSKSYTKLCRFSLEIFDVGLIEISARNMGLETC 1705

Qy   1021 VRTENAI-----NSPVG---------------ILETSRVAVSKVVAGTKEFWSEVSLNDF 1060
Db   1706 VSMASAIFDVILTYPLGSISFPVVCARVFGALAMEVGSNYCFKKTLGKPEILGEDAFRRT 1765

Qy   1061 TTFVLRNKV--------------------LIGIFVASLGAAPIAW----KYRRG----IA 1092
Db   1766 VSMVLSAKAFDKLSVDAKGLVRCSGIIPGVLRIVFSSLFSDADAWVGYSKHEVSSLPMIT 1825

Qy   1093 ANARRYA---------------GSSYETLSSLSSQ------------AAGGLRGLTSSTV 1125
Db   1826 YSAKRLSFVLEQTKEACIAFIVDSSKRILASLTSEIHESLQNSELTSGAKSLYGRVSSIL 1885

Qy   1126 SGGSLVVRRGFSSAVTVTRA-TVAKRQVPLALLSF----------STSYAISGCSMLGIW 1174
Db   1886 SASNGMVNKAGNAVAAVKAAGSGLSKRVKRIRPSFHLTRCVNEQTSEYYSASDDSELD-- 1943

Qy   1175 AHALPRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE---VVWRGKSVSSL-LLPI 1230
Db   1944 -ESL---------LSDTPGLRGN-------GKCNDLFRVVSGRVLVYRRRAVEALKRLVL 1986

Qy   1231 TLG-----------VSLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRY--YRDFD---YD 1274
Db   1987 HVGECLYSRIADSFVRLFVR--TNTAVGKCESLAVDCATKITDLVVGYITWRNLHFGAYS 2044

Qy   1275 EGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQ-- 1332
Db   2045 TTPGLSGGSRVSLSRLLDDGLTC------VIQNFVTAWLTIFNETFYSSLAHRMRTASIN 2098

Qy   1333 ----------------YTYVE--EEVAP------------------SAAVAE------- 1348
Db   2099 VLLEFLKNLPVCLVSIYTFIECRSRIFPLKEFLWGFCIFVWNIIKELSVCIYENLPVCVT 2158

Qy   1349 --------RQGDPSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEVDGKGLLPESV 1400
Db   2159 RLCNRAFRRSSNEDDERTLQVITEAENRIGFLREVLAEMARYRASLEIPVPGS---PNSP 2215

Qy   1401 VGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEV 1460
Db   2216 LFPPTYGEIEELGDGSSDSARVIGVDNGSDSMSTEGEAELDSDGDSDDSIPLDRDFEQIG 2275

Qy   1461 PLATVSGATPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKA 1520
Db   2276 GL----GGSGNVD------VFRLGLRFV--WRMSRNIFNLRAVKFSYGLALQCIFNGRST 2323

Qy   1521 G------------------KQLCMFRTCSCGVQLDVYN-----EATIA-----TRF--- 1548
Db   2324 GQFISRLLFSFGGCLPFEFFRHVCAVRWLSRSLALHVRNASICDEETVEWLFNFSRFLYD 2383

Qy   1549 ----SNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWP-RALEDILTAIKYPSVFD 1603
Db   2384 SRYLSLVFSELERVEFYLSYLFSLPKRCFTHNRSVQVCVRAPVRVVRDV----------- 2432

Qy   1604 HCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCRKGGKVMVINVASGDY 1663
Db   2433 ----------GLV------EDGRKSLSEVL--NSLDSLKVS---NRKGKSAVIESDSSD- 2470

Qy   1664 FLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGV--GRMLQLA-GGVSDEKSPGVPNQ 1720
Db   2471 ------------SESGSSSGSIS---------FGSINEEVLKTANGSVVSEKA-----S 2503

Qy   1721 QPQSQGATRTITPKSGGK----------ALSEGSGREVKGRSTYSIWCEQDYVRKCEWLR 1770
Db   2504 LMKHAAALSGLEPDLGGKKDLPIIRRNFTRGESSGTKSVG------------SLCEYLT 2550

Qy   1771 ADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEG 1830
Db   2551 TLNSSFGSVPDLYPVARDIT----------YKK-----MTNAMREFYYSQKVTLYELHG 2594

Qy   1831 VLKVPNQVYESL--PGFHVYKSGTDLIFHSTQDGLRVRDLPY-VLIAEKGI--FTKGKDV 1885
Db   2595 KL---SSYWDELKTAGFDRKLAKMDL-----DDDVGVVDFNLKVIVGRHGTRPFRNTVNS 2646

Qy   1886 DAVVALGDNLFVCD-----DILVFHDAINLIGALKVARCGMVGE--SFKSFEYK--CYNA 1936
Db   2647 HEFMFCDDGLVLCPGPKKRDFALISSQTQFVAANSFLR-AVDGKDLTFTNSEHSLIVYEA 2705

Qy   1937 PPGGGKTTTLVDEFVK--SPNSTATITANVGSSEDINMAVKKRDPN--------LEGLNS 1986
Db   2706 PPGGGKTHSLVTYFADYCTKVSCLVVTANKNSQAEISQRISRELMDRKMLAKHVIKTAGR 2765

Qy   1987 ATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINRE 2046
Db   2766 VFTVDSYLMNHI-RLKTDLLFVDECFMVHAGAIGAVVEFTSCKAVVFFGDSRQIHYIHRN 2824

Qy   2047 KV---FRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKP 2103
Db   2825 DLGVSLLHDIDAFI-EPQHRIYGEVSYRCPWDICEWLS--------KFYPRHVATANTGS 2875

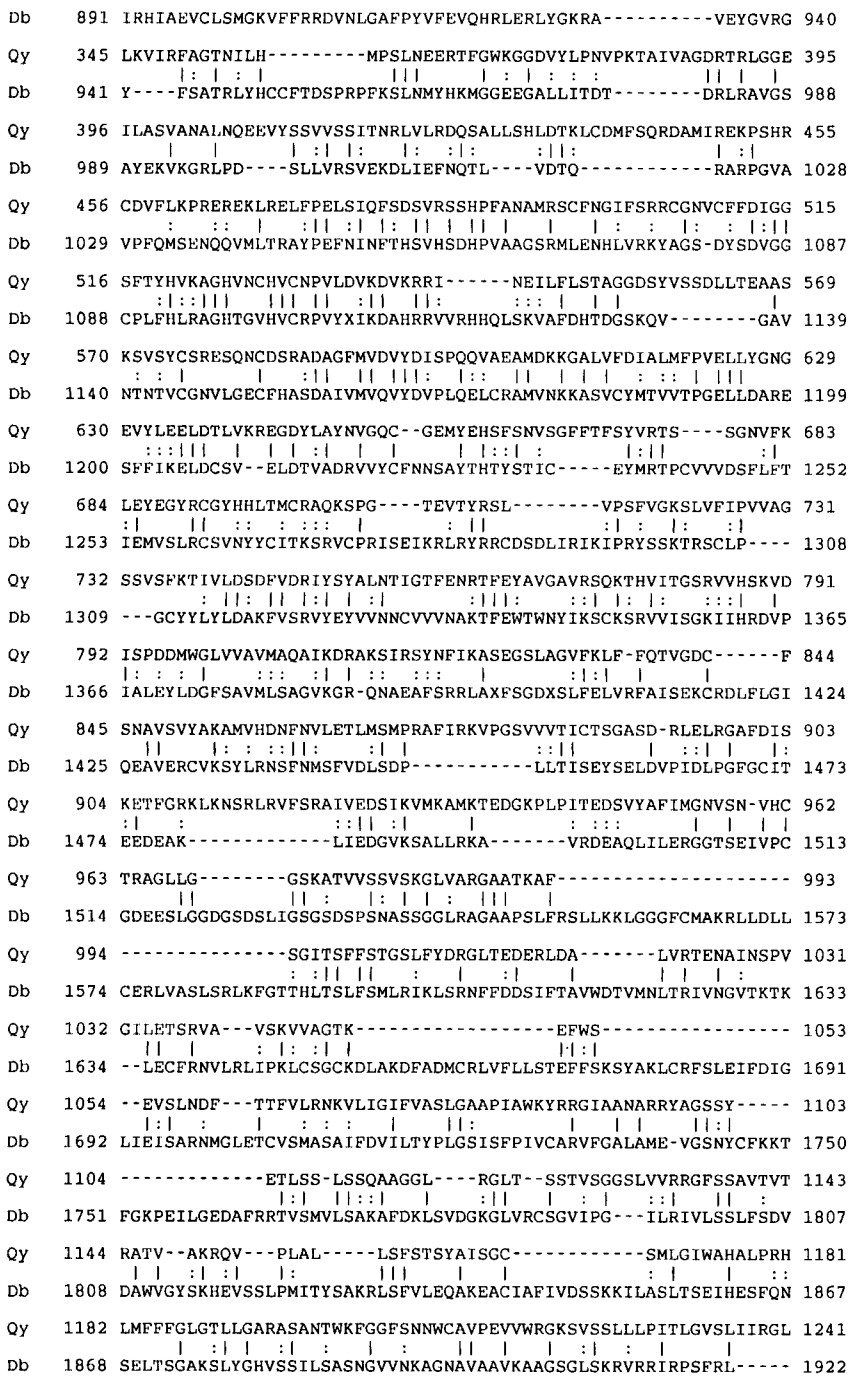
Qy   2104 VVR-SLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSL----KGKGKETPVM-TVHEA 2157
Db   2876 VGKSSVSIEVINGCDDVPYDSSAKYIVYTQAEKNELQKHLGRLTVGRTKAVPVVNTVHEV 2935

Qy   2158 QGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDA 2215
Db   2936 QGETYKKVRLVRCKYQEDTPFCSDNHVVVALTRHVDSLTYSVLNSRRYDKTASSIDEA 2993
```

RESULT 11
O10467
ID O10467 PRELIMINARY; PRT; 3115 AA.
AC O10467;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE 347-KDA POLYPROTEIN.
OS Citrus tristeza virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=12162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SY568;
RA Yang Z.N., Mathews D.W., Williams A., Dodds J.A., Mirkov T.E.;
RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF001623; AAB57702.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
KW Polyprotein.
SQ SEQUENCE 3115 AA; 347072 MW; B406EA0FB38E73F9 CRC64;

Query Match 5.9%; Score 678.5; DB 12; Length 3115;
Best Local Similarity 20.5%; Pred. No. 1.6e-37;
Matches 482; Conservative 308; Mismatches 880; Indels 681; Gaps 99;

```
Qy    298 MRERAFVCTTKGWCWFNNERLRG------------EIYRRRCFSSSFSIGFLMHLGFRS 344
```





```
Qy   2091 CYPEKVVSGKDKPVVR-SLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSL----KGK 2145
Db   2863 FYPRHVATANTGSIGKSSVSIEIINGCDDVPYDSSAKYIVYTQAEKNELQKHLGRLTVGR 2922

Qy   2146 GKETPVM-TVHEAQGKTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKL 2204
Db   2923 TKAVPIVNTVHEVQGETYKKVRLVRCKYQEDTPFCSDNHVVVALTRHVDSLTYSVLNSRR 2982

Qy   2205 DDKVGTYISDA 2215
Db   2983 YDKTASSIDEA 2993


RESULT 12
P87587
ID   P87587      PRELIMINARY;      PRT;  3107 AA.
AC   P87587;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   P346.
OS   Citrus tristeza virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=12162;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VT;
RX   MEDLINE=96406950; PubMed=8811037;
RA   Mawassi M., Mietkiewska E., Gofman R., Yang G., Bar-Joseph M.;
RT   "Unusual sequence relationships between two isolates of citrus
RT   tristeza virus.";
RL   J. Gen. Virol. 77:2359-2364(1996).
DR   EMBL; U56902; AAB38755.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
SQ   SEQUENCE   3107 AA;  346659 MW;  02B748F693BE89CB CRC64;


  Query Match             5.7%;  Score 661.5;  DB 12;  Length 3107;
  Best Local Similarity   20.4%;  Pred. No. 2.5e-36;
  Matches  480;  Conservative  337;  Mismatches  882;  Indels  655;  Gaps  102;

Qy    298 MRERAFVCTTKGWCWFNNERLRG------------EIYRRRCFSSSFSIGFLMHLGFRS 344
Db    885 IRHIAEVCLSMGKVFFRRDVNLGAFPYVFEVQHRLERLYGKRA----------VEYGVRG 934

Qy    345 LKVIRFAGTNILH---------MPSLNEERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGE 395
Db    935 Y----FSATRLYHCCFTDSPRPFKSLNKYHKMGGEEGALLITDT--------DRLRAVGS 982

Qy    396 ILASVANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQRDAMIREKPSHR 455
Db    983 AYEKVKGRLPD----SLLVRSVEKDLIEFNQTL----VDTQ-----------RARPGVA 1022

Qy    456 CDVFLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCFFDIGG 515
Db   1023 VPFQMSENQQVMLTRAYPEFNINFTHSVHSDHPVAAGSRMLENHLVRKYAGS-DYSDVGG 1081

Qy    516 SFTYHVKAGHVNCHVCNPVLDVKDVKRRI------NEILFLSTAGGDSYVSSDLLTEAAS 569
Db   1082 CPLFHLRAGHTGVHVCRPVYDIKDAHRRVVRHHQLSKVAFDHTDGSKQV--------GAV 1133

Qy    570 KSVSYCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNG 629
Db   1134 NTNTVCGNVLGECFHASDAIVMVQVYDVPLRELCRAMVNKKASVCYMTVVTPGELLDACE 1193

Qy    630 EVYLEELDTLVKREGDYLAYNVGQC--GEMYEHSFSNVSGFFTFSYVRTS----SGNVFK 683
Db   1194 SFFIKELDCSV--ELDRSPNRVIYCFNNSRYTHTYSTIC-----EYMRTPCVVVDGFLFT 1246

Qy    684 LEYEGYRCGYHHLTMCRAQKSPG----TEVTYRSL--------VPSFVGKSLVFIPVVAG 731
Db   1247 IEMVSLRCRVNYYCITKSRVCPRISEIKRLRYRRCDSDLIRIKIPRYSNKTRSCLP---- 1302

Qy    732 SSVSFKTIVLDSDFVDRIYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVD 791
Db   1303 ---GCYYLYLDAKFVSRVYEYVVNNCVVVNAKTFEWTWNYIKSCKSRVVISGKIIHRDVP 1359

Qy    792 ISPDDMWGLVVAVMAQAIKDRAKSIRSYNFIKASEGSLAGVFKL----FFQTVGDCF--- 844
Db   1360 IGLEYLDGFSAVMLSAGVKGRQNAETFSRRLRAFSGD-PSLFELVRFAISEKCRDLFVGV 1418

Qy    845 SNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISK 904
Db   1419 QEAVERYVKSYLRNSFNISFADLSDP----------LLTI-----SEYSELDVPIDL-- 1460

Qy    905 ETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTEDGKPLPITE----------------- 946
Db   1461 PGFGCITEGDEAK-----LIEDGVKSALLRKAFVTRPQLILERGGTSEIVPCGDDDGLGS 1515

Qy    947 ---DSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGL-----VAR----------G 987
Db   1516 DGSDSLLG--SGSDSSSKASPGGLRAG-RPSLFRSLLKKLGGGFCMAKRLLDLLCERLVG 1572

Qy    988 AATKAFSG---ITSFFSTGSLFYDRGLTEDERLDA------LVRTENAINSPVGILETS 1037
Db   1573 SLSRFKFGTTHLTSFFSMLHAKLSRNFFDDSIFTAVWDVVMNLIRIVNGVTRTK--LECF 1630

Qy   1038 RVA---VSKVVAGTK----------------EFWS----------------EVSL 1057
Db   1631 RNVLRLIPTLCSGCKDFAKDFANMCRLVLPFSTEFFSKSYTPLCRFSLDVFDVGLIEISA 1690

Qy   1058 NDF---TTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSY---------- 1103
Db   1691 RNMGLETCVSMASAIFNVVLTYPLGSISFPVVCARVFGALAME-VGSNYCFKKSFGKPEI 1749

Qy   1104 -------ETLSSLSSQAA----------------GGLRGLTSSTVSGGSLVVRRGFS 1137
Db   1750 LGEDAFRRTVSMVLSAKAFDKLSIDSKGLVRCSGVIPGVLRIVFSSLFSDADAWV--GYS 1807

Qy   1138 ----SAVTVTR----------ATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLM 1183
Db   1808 KHEVSSLPITRVARKSYLSRWSRLRKRASPL----YDSSKRI--LASLTSEIHESFQNLE 1861

Qy   1184 FFFGLGTLLGARASANTWKFGGFSNNWCAVPEVVWRGKSVSSLLLPITLGVSL------- 1236
Db   1862 LTNGAKSLYGHVSSVLSASNSVVSKAGNAVAAVKAAGSGLSKRVRRIRPSFRLARCVNEQ 1921

Qy   1237 ------------IIRGLLNDTIPQL-------AYVPPVEGR------------------ 1258
Db   1922 TSEYNSASDESELDESLLSDT-PGLRGNGKCKALFRVISGRVSVCCRRAVEALKRLVLHV 1980

Qy   1259 ------NVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGSTSSVSSYDVVTNVRDVG 1312
Db   1981 GDCLYTRIMDSFVRLFVETNVAVG------KCESFVVDSTRKISDLVNGYITWRNLHSGA 2034

Qy   1313 ISTNGEVTGEEETHSPRSVQ----------YTYVEE-EVAPSAAVAERQGDPSGSGTAD 1360
Db   2035 YTTTPGLSGGSRSSLSRFLDGGLTVCNSKYCYRVVDRFSMRPLLSLVHRM------RTAS 2088

Qy   1361 AMAFVESVKK------GVDDVFHQQSS--------------GETAREVEVDGKGLLPES 1399
Db   2089 INVLLEFLKNLPVCLVSIYTLIECRSRIFPLKEFLWGFCVFVWNIVKELSICIYGSLPVC 2148

Qy   1400 VVGEAPTQERGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSE--VHAQKEVK 1457
Db   2149 V-----TRLCNRA-----FRRGSNEGD--------VGTPQV----ITEAENRIEFLREVL 2186

Qy   1458 QEVPLATVSGATPIVDEKPAPSV------------TTRGVKIIDKGKAVAHVAEKKQVQ 1504
Db   2187 AEMARYRASLEIPVPGSPDSPFVPPTYGEIEELCNDSSDSARLIGVDNGLDSVPTEGEAE 2246

Qy   1505 VEQ--------PKQRSLTINEGKAGKQLCMFRTCSCGVQLDVYNEATIATRFSNAFTFVD 1556
```

```
Db   2247 LDSDGDSDDSIPLDRDFEQIGGLGGSG----------NVDVFR---LGLKF--VWRMLR 2290

Qy   1557 NLKGRSAVFFS-KLGEGYTYNGGS---------------------HVSS-GWPRALEDI 1592
Db   2291 NLFNLRAVKFSYGLALQCIFNGGSMGQFTSRLLFSFGNCLPFELFRHVCSVRW--LSKSL 2348

Qy   1593 LTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANF--------- 1643
Db   2349 AIRVRNARICDEETVERFFSFSRILY----DSRYLSD----VFSVLERVDFYLSYLFSLP 2400

Qy   1644 --------STKCRKGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRI--------- 1686
Db   2401 KRCLTRNRSVRVTRRAPVRVIR----DMGLVEDG-KKSLTEVLNSLDSLKVSDRKGKSAL 2455

Qy   1687 --------------SLTFRATRRVFGVGRMLQ-LAGGVSDEKSPGVPNQQPQSQGATRT 1730
Db   2456 IESDGSDSESGHSFGSISFGSVKE-----EVLEAVNSGVVSEKAPLMKHADALSR----- 2505

Qy   1731 ITPKSGGK----------ALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEP 1780
Db   2506 LEPDLSGKKDLPIVRRSFTRGESTGTKPAG------------SLCVYLTTLNSSFGSVP 2552

Qy   1781 DYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMA-------RNIAVTTAEGVLK 1833
Db   2553 DLYPVARDI----------------------TYKKMTNAMPKFYYSQKVTLYELHGKL- 2588

Qy   1834 VPNQVYESLP--GFHVYKSGTDLIFHSTQDGLRVRDLPY-VLIAEKGI--FTKGKDVDAV 1888
Db   2589 --SSYWDELKTVGFNRKLAKMDL-----DDDVDVVDFNLKVIVGRHGTRPFRNTVNSHEF 2641

Qy   1889 VALGDNLFVCD-----DILVFHDAINLIGALKVARCGMVGE--SFKSFEYK--CYNAPPG 1939
Db   2642 MFCDDGLVLCPGPKKRDFALISSQTQFVAANSFLR-AVDGKDLTFTNSEHSLIVYEARPG 2700

Qy   1940 GGKTTTLVDEFVK--SPNSTATITANVGSSEDINMAVKKRDPNLEGL--------NSATT 1989
Db   2701 GGKTHSLVTYFARYGTKVSCLVVTANKNCQAEISQRISRELMDRKMLAKHVVKTAGRVFT 2760

Qy   1990 VNSRVVNFIVRGMYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKV- 2048
Db   2761 VDSYLMNHI-RLKTDLLFVDVCFMVHAGAIGAVVEFTSRKAVVFFGDSRQIHYIHRNDFG 2819

Qy   2049 --FRMDCAVFVPKKESVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGKDKPVVR 2106
Db   2820 VSLLHDIDAFI-EPQHRIYGEVSYRCPWDICEWLS--------KFYPRHVATANTGSIGR 2870

Qy   2107 -SLSKRPIGTTDDVAEINADVYLCMTQLEKSDMKRSLKGK-----GKETPVM-TVHEAQG 2159
Db   2871 SSVSIEIINGCDDVPYDSSAKYIVYTQAEKNELQKHL-GRLTVAGTKAVPIVNTVHEVQG 2929

Qy   2160 KTFSDVVLFRTKKADDSLFTKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQS 2219
Db   2930 ETYKKVRLVRCKYQEDTPSCSDNHVVVALTRHVDSLTYSVLNSRRYDKTASSIDEA--RE 2987

Qy   2220 VSDALLHTFAPAGC 2233
Db   2988 IFDKFRSTNHSHGC 3001


RESULT 13
Q83044
ID   Q83044      PRELIMINARY;      PRT;  1873 AA.
AC   Q83044;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   METHYLTRANSFERASE.
OS   Lettuce infectious yellows virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC   Closterovirus.
OX   NCBI_TaxID=31713;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=92;
RA   Klaassen V.A., Boeshore M., Koonin E.V., Tongyan T., Falk B.W.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U15440; AAA61797.1; -.
DR   INTERPRO; IPR000606; -.
DR   INTERPRO; IPR002588; -.
DR   PFAM; PF01443; Viral_helicase1; 1.
DR   PFAM; PF01660; Vmethyltransf; 1.
KW   Transferase; Methyltransferase.
SQ   SEQUENCE   1873 AA;  217264 MW;  55A771D8357761F0 CRC64;


  Query Match             5.6%;  Score 643;  DB 12;  Length 1873;
  Best Local Similarity   18.4%;  Pred. No. 1.8e-35;
  Matches  406;  Conservative  323;  Mismatches  733;  Indels  748;  Gaps   83;

Qy    166 VKPTAVGVHVPLPKKQEALEPAQSVPQQSLEEKAALTFGLFFSKGGGDESDAVILRKGKL 225
Db    259 VKPRVVSVN-----KAREVSSSVEVKLQRYEREGEV--------------FVVSKAGKE 298

Qy    226 FNRALNVPIDVKNTFVWAKIWDEASRRRGYFYVKDRAVKFFPIVRGRATIEDFIVNTAPG 285
Db    299 DTRVLNDDNAIRNLF-------NATLNGGKY-------KLHP----------------- 326

Qy    286 CDVALPRIELWSMRERAFVCTTKGWCWFN-----NERLRGEIYRRRCFSSSFSIGFLMHL 340
Db    327 --------EAKSATGKRFKYYKDGFCWLDVFADANRRI-PEWVKPHCLLTG---SVLMSC 374

Qy    341 G---FRSLKVIRFAGTNILHMPSLNEERTFGWKGGDVYLPNVPKTAIVAGDRTRLGGEIL 397
Db    375 GLWDFAKRKMVS-VSHGLLH----------------YDRKLERSSARAGVRDFVG---- 412

Qy    398 ASVANALNQEEVYSSV---VSSITNRLV----LRDQSALLSHLDTKLCDMFSQRDAMIRE 450
Db    413 --ASNEAVQREDFRDLDLRVEEFAERVLETANLRSDNRLIDNILTRASDYINKKSKESKE 470

Qy    451 KPSHRCDVFLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCF 510
Db    471 LDINVC---LSMDEKKMITNLFPDIQMSFNQKSYSNHGVFNAMRACENFYFSRKFKNSDY 527

Qy    511 FDIGGSFTYHVKAGHVNCHVCNPVLDVKDVKRRINEILFLSTAGGDSYVSSDLLTEAASK 570
Db    528 IDAGGDVVSTLRSKNHNVHICSPRLDLKDAARHIQRATVIDGLKG----------YGE 575

Qy    571 SVSYCSRESQNCDSRADAGFMVDVYDISPQQVAEAMDKKGALVFDIALMFPVELLYGNGE 630
Db    576 TISFCTNKTEDCAVNRDIIIAVEVYDMTLRDMAKAMLSHGSRKFEFSCIIPPEIFTKECN 635

Qy    631 VYLEELDTLVKREGDYLAYNVGQCGEMYEHSFSNVSGFFTFSYVRTSSGNVFKLEYEGYR 690
Db    636 VELYEGRLKVTRIGDNVEYYYGSNGETFSHSCQTLKDILSVQ-VFQFGGRVFKKTLEHSR 694

Qy    691 CGYHHLTMCRAQK-SPGTEVTYRSLVPSFVGKSLVFIPVVAGSSVSFKTIVL-DSDFVDR 748
Db    695 GQLHFFSICICEKIEPGSVKLKTYYQRSELDKVTLRIPVKDSFGVVTHYIIKEDREFVSS 754

Qy    749 IYSYALNTIGTFENRTFEYAVGAVRSQKTHVITGSRVVHSKVDISPDDMWGLVVAVMAQA 808
Db    755 MIEYVANTGIKIDDKMVEWTYSQYRAKKTVTIKSGKVTQKETRIRKELIPGFIAIIMSEG 814

Qy    809 IKDRAKSIRSYNFIKASEGS--------------LAGVFKLFFQTVGDCFSNAVSVYAKA 854
Db    815 IRAREKTHYLAKMLYTSHYKPSIVNIIFRLIMHFLGGTKRFIYESLVDCLKFLT------ 868

Qy    855 MVHDNFNVLETLMSMPRAFIRKVPGSVVVTICTSGASDRLELRGAFDISKETFGRKLKNS 914
Db    869 ----NSDYIDTIVN--------------------TESRIE-----DLDK---------- 888

Qy    915 RLRVFSRAIVEDSIKVMKAMKTEDGKPLP-ITEDSVYAFIMGNVSNVHCTRAGLLGGSKA 973
Db    889 ------WFVFEQNVTI-----TTDAEDQPSILEQSVKTFL------------------- 917
```

```
Qy    974 TVVSSVSKGLVARGAAT----KAFSGITSFFSTGSLFYDRGLTEDERLDALVRTENAINS 1029
Db    918 ---SKYSEDVYERGDRSEFELEEFS----------------SEQDLVDQL------LNS 951

Qy   1030 PVGILETSRVAVSKVVAGTKE---FWSEVSLNDFTTFVLRNKVLIGIFVASLGAAPIAWK 1086
Db    952 G--------------GGSKEDEWFYS------YIFSIVRTKCTLS----------LAWR 980

Qy   1087 YRRGI-------AANARRYAGSSYETLSSLSSQAAGGLRGLTSSTVSGGSLVVRRGFSSA 1139
Db    981 WTNRIFNFITSCRAKGKDFVKYIYEIIKIIINKP---WEALCHYLVQGYKTVLK------ 1031

Qy   1140 VTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHALPRHLMFFFGLGTLLGARASAN 1199
Db   1032 ------------IPMTCVD-SVNY---------LKRKSLEKILSYF-------SHERADD 1062

Qy   1200 TWKFGGFSNNW------------------CAV----PEVVWRGKSVSSLLLPITLGVSL 1236
Db   1063 EYIFPGYDSDFSEESGDYGDENTELDSRSRCTVIKSKVEMLFRN---------------- 1106

Qy   1237 IIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQHEAVPGDDNDGST 1296
Db   1107 -IKGHIEKFLQRCGII------DQYNKIYHWFKSI----------------------L 1135

Qy   1297 SSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQGDPSGS 1356
Db   1136 TDVDCFNSVLNF---------------ITHT------------GCDALLAVLTGHFSIA 1167

Qy   1357 GTADAMAFVESV--KKGVDDVFHQQSSGETAREVEVDGKGLLPESVVGEAPTQERGRAAD 1414
Db   1168 SC--ALKFMTDVVIEKNLSEKHHTTT----------------------------KLIA 1195

Qy   1415 GNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGATPIVDE 1474
Db   1196 GNLTQCIYKLDFIHPIWIPI----RWGVKEI--SKIHIKKK---------------LLNW 1234

Qy   1475 KPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQR---SLTINEGKAGKQLCMFRTCS 1531
Db   1235 RSTKEVTNESI-----CKDLVHKSYIKYFDIKKAVWLLWFSLLI--------IFLHPTLG 1281

Qy   1532 CGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTY---------------N 1576
Db   1282 FFILFSIYPAIEIKRYYNNVVCF-NNIKMSYPHVIDRLGGTYNFAKLKKAVREKFSENKS 1340

Qy   1577 GGSHVSSGWPRALEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADD---EECYPSDNPIL 1633
Db   1341 QGVKISEIPAEKDEDVMPPLEE--------VETPKRSVGRPSSKQDDIEDECNIKEDPTE 1392

Qy   1634 TVNL--------------VGKANF--------STKCRKGGKVMVINVASGDYFLMPCGFQ 1671
Db   1393 SLNFSNIEATSNRYDTLRVGNSKLCKFLRFYPKSKC-------YANIQTGDLI------- 1438

Qy   1672 RTHLHSVNSIDEGRISLTFRATRRVFGV----GRMLQLAGGVSDEKSPGVPNQQPQSQGA 1727
Db   1439 RDSIEEFYHLERSKLDIEISKMQRVVEIMNDTGRMLDNVRRMIDDRSMYV---------- 1488

Qy   1728 TRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEWLRADNPVMALEPDYTPMTF 1787
Db   1489 ------------------TLDGVSWYRLGCKD-------------------------- 1502

Qy   1788 EVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIAVTTAEGVLKVPNQVYESLPGFHV 1847
Db   1503 ---KNPAKED---------------------------------------------- 1509

Qy   1848 YKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKDVDAVVALGDNLFVCDDILVFHDA 1907
Db   1510 FKAIFDHNFNIMEGNAKVKEFAVSSDEWRGMYSNER-CRAI----EQLFNDNNEMVRRPD 1564

Qy   1908 INLIGALKVARCGMVGESFKSFEYKCYNAPPGGGKTTTLVDEFVKSPNSTATITA----N 1963
Db   1565 VN----------GL----------KFYNKPPGAGKTTTIAKLMSKDLKNKVKCLALSYTK 1604

Qy   1964 VGSSEDINM----AVKKRDPNLEGLNSATTVNSRVVNFIVRGMYKRVLVDEVHMMHQGLL 2019
Db   1605 VGRLELIDKLKKDGIEKPEKYVKTYDSFLMNNDNILEIV------NLYCDEVFMMHAGHF 1658

Qy   2020 QLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVP--------KKESVVYTSKSYR 2071
Db   1659 LTLLTKIAYQNGYCYGDVNQIPFINRDP--------YTPAYLSREFFRKQDLNYDTYTYR 1710

Qy   2072 CPLDVCYLLSSMT-----------VRGTEKCYPEKVVSGKDKPVVRSLSKRPIGTTDDV- 2119
Db   1711 CPLDTCYLLSNLKDEMGNIIYAGGVKNVNEVY----------PTIRSLNLFGINVVGEVP 1760

Qy   2120 AEINADVYLCMTQLEKSDMKRSLKGKGK-ETPVMTVHEAQGKTFSDVVLFRTKKADDSLF 2178
Db   1761 VEYNAK-YLTFTQDEKLNLQRHIDSQGGCRNAVSTVNEAQGCTFSEVNLVRLVQFDNPVM 1819

Qy   2179 TKQPHILVGLSRHTRSLVYAALSSKLDDKVGTYISDASPQSVSDALLHTF 2228
Db   1820 SDINQFVVAISRHTTTFKYFTPHSRLNDRVSNAIS--SLQSVSDFVLKDY 1867
```

RESULT 14
Q9JGK3
ID Q9JGK3 PRELIMINARY; PRT; 1334 AA.
AC Q9JGK3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE 152-KDA PROTEIN.
OS soil-borne wheat mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
OX NCBI_TaxID=28375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAPANESE;
RA Shirako Y., Suzuki N., French R.C.;
RT "Similarity and divergence among viruses in the genus Furovirus.";
RL Virology 270:201-207(2000).
DR EMBL; AB033689; BAA94796.1; -.
SQ SEQUENCE 1334 AA; 151950 MW; 288761E57551D0A7 CRC64;

Query Match 4.2%; Score 483.5; DB 12; Length 1334;
Best Local Similarity 19.5%; Pred. No. 1.7e-24;
Matches 370; Conservative 237; Mismatches 607; Indels 683; Gaps 75;

```
Qy    398 ASVANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQR--DAMIREKPSHR 455
Db      6 SNILGIINEEEVVRAAIATSATR--------TNSELHRTVCEQIREQFLDTVENQKQKKK 57

Qy    456 CDV--FLKPREREKLRELFPELSIQFSDSVRSSHPFANAMRSCFNG-IFSRRCGNVCFFD 512
Db     58 IDVRRDLTQEQLQTLNELYPERHIVTSGCERGTHSFAAASRKIETDLLLSRMPKNANVYD 117

Qy    513 IGGSFTYHVKAGHV-NCHVCNPVLDVKDVKRRINEIL----FLSTAGG------------ 555
Db    118 IGGNWATHLKRKDTRKVHCCCPVLDFRDAQRKTTRWLSVEKFLSEKDSISEHCGQKVLEI 177

Qy    556 ---DSYVSSDLLTEA-ASKSVS---YCSRESQNCDSRADAGF---MVDVYDISPQQVAEA 605
Db    178 QEDEDRISNNLRKGAVAPEDLTGRWYCENRFEDCVYRAEKAYGMAIHSIYDISLDDLVNA 237

Qy    606 MDKKGALVFDIALMFPVELLYGNGEVYLEELDTLVKREG---DYLAYNVGQCGEMYEHSF 662
Db    238 MEEKRIKYLVGTFLFSVELFLGKKRGELTSVDGFFEIDGGNVKYGFYDDTNCG--YRHDL 295

Qy    663 SNVSGFFTFSYVRTSSGNVFKLEYEGYRCGYHHLTMCRAQKSPGTEVTYRSLVPSFVGKS 722
Db    296 QQLMEYLTKTFVKAKGGSVFYLELTEQRGDVMFFSLTDA-----TEARMHGVVED---ES 347

Qy    723 LVFIPVVAGSSV--------------SFKTIVLDSDFVDRIYSYA-------LN--TIGT 759
Db    348 FKCIPIESKDSVVFPLFELNKKSDELEFTEVLLPKSFVRRTIEYTARLKTNQLNPETVNS 407
```

```
Qy    760 FENRTFEYAVGAVRSQKTHVITGS-RVVHSKVDISPDDMWGLVVAVMAQAIKDRAKSIRS 818
Db    408 YLTST----------NNTVIIGGSAKKTVEKVDAMLIPQITTTLIVWTELMNARQKKV-- 455

Qy    819 YNFIKASEGSLAGVFKLFFQTVGDCFSNAVSVYAKAMVHDNFNVLETLMSMPRAFIRKVP 878
Db    456 ------------LDRLRIQMKDD-------VGFMALAHASFSKMFGKVS----FYQKAL 491

Qy    879 GSVVVTICTSGASDRLELRGAFDISKETFGRKLKNSRLRVFSRAIVEDSIKVMKAMKTED 938
Db    492 RSYANWISYCHGTDAIE--------------FKNVPLYV----EVKDRVKMWKQHAPNH 532

Qy    939 GKPLPITEDSVYAFIMGNVSNVHCTRAGLLGGSKATVVSSVSKGLVARGAATKAFSGITS 998
Db    533 G----------FSFDLEDL---------------------------------------- 541

Qy    999 FFSTGSLFYDRGLTEDERLDALVRTENAINSPVGILETSRVAVSKVVAGTKEFWSEVSLN 1058
Db    542 --------------DEKIQLYKETER-----------ERKAISEFVVSDK-------LG 568

Qy   1059 DFTTFVLRNKVLIGIFVASLGAAPIAWKYRRGIAANARRYAGSSYETLSSLSSQAAGGLR 1118
Db    569 DLSAAVCEKE-----------AREVSVEYR------DRR-------------------R 591

Qy   1119 GLTSSTVSGGSLVVRRGFSSAVTVTRATVAKRQVPLALLSFSTSYAISGCSMLGIWAHAL 1178
Db    592 KVEATDLRNGEV------------------------------------------------ 603

Qy   1179 PRHLMFFFGLGTLLGARASANTWKFGGFSNNWCAVPE----VVWRGKSVSSLLLPITLGV 1234
Db    604 ----------------------FTNFVDEWCNKEDHFNHVHIEVENKYSLLVKIVIAV 639

Qy   1235 SLIIRGLLNDTIPQLAYVPPVEGRNVYDETLRYYRDFDYDEGAGPSGTQH--EAVPGDDN 1292
Db    640 -------------WEAIVPPMQFAPVYID----------DEERDPSDAESVVELDLSETD 676

Qy   1293 DGSTSSVSSYDVVTNVRDVGISTNGEVTGEEETHSPRSVQYTYVEEEVAPSAAVAERQGD 1352
Db    677 DGSGKG-----------DYAEKPGSVEKAEKPIEARVSLEKGKEKVEVEDSAI---RRG- 721

Qy   1353 PSGSGTADAMAFVESVKKGVDDVFHQQSSGETAREVEV--DGKGLLPESVVGE--APTQE 1408
Db    722 -SGSSSPDS--------------------ETSMEDELVSDGESLFGRSCQATIYSSTSL 759

Qy   1409 RGRAADGNTAQTAVNEGDREPVQSSLVSSPQADIPKVTQSEVHAQKEVKQEVPLATVSGA 1468
Db    760 YGSSEDDTSVGSASSE-----IDSRVLEAVEAAIDEM-------EKEVSAEVLAAVV--- 804

Qy   1469 TPIVDEKPAPSVTTRGVKIIDKGKAVAHVAEKKQVQVEQPKQRSLTINEGKAGKQLCMFR 1528
Db    805 --VVTESSEPEVSEVVV------ENAATVECKEQVVVDTTDSES---DEERAK------- 846

Qy   1529 TCSCGVQLDVYNEATIATRFSNAFTFVDNLKGRSAVFFSKLGEGYTYNGGSHVSSGWPRA 1588
Db    847 --SWG---SMAEEESDDTFYMNTM-LISNKVQRSSL--PKQPDFNKYN------TVQQKA 892

Qy   1589 LEDILTAIKYPSVFDHCLVQKYKMGGGVPFHADDEECYPSDNPILTVNLVGKANFSTKCR 1648
Db    893 KQEYLWYLRCKMISDRTTLRS----------------------IIEDHLNG--------- 921

Qy   1649 KGGKVMVINVASGDYFLMPCGFQRTHLHSVNSIDEGRISLTFRATRRVFGVGRMLQLAGG 1708
Db    922 -------------------------LYHNGNCDLPKNSCFLDYTKSVGGEWLYAK---- 951

Qy   1709 VSDEKSPGVPNQQPQSQGATRTITPKSGGKALSEGSGREVKGRSTYSIWCEQDYVRKCEW 1768
Db    952 ----------------------PLKAGHCYGVGFSLNSKGK----------IGKCEL 976

Qy   1769 LR----ADNPVMALEPDYTPMTFEVVKTGTSEDAVVEYLKYLAIGIERTYRALLMARNIA 1824
Db    977 LKLMWDVDARGQPLERPFNTKMFQFV--------LLNDLTFLM---------------- 1011

Qy   1825 VTTAEGVLKVPNQVYESLPGFHVYKSGTDLIFHSTQDGLRVRDLPYVLIAEKGIFTKGKD 1884
```

```
Db   1012 ----------NEM---------------IIFRNLQDTLQRKER----------TKQAS 1034

Qy   1885 V---DAVVALGDNLFVCDDILVFHDAINLIGALKVARCGMVGESFKSFEYKCYNAPPGGG 1941
Db   1035 IVLKDGVPGCGKSTWILNNANYIKDVVISVG--KEAKEDLKEKFMK--KYKCTESELGRI 1090

Qy   1942 KTTTLVDEFVKSPNSTATITANVGSSEDINMAVKKRDPNLEGLNSATTVNSRVVNFIVRG 2001
Db   1091 RT---VDSYL--------------MHDCGKKLR----------ATTVH---------- 1111

Qy   2002 MYKRVLVDEVHMMHQGLLQLGVFATGASEGLFFGDINQIPFINREKVFRMDCAVFVPKKE 2061
Db   1112 ------FDEALMTHAGAVYFCADLLGARKVICQGDSQQIPFVNRVESIKLQFAKLVIDKT 1165

Qy   2062 SVVYTSKSYRCPLDVCYLLSSMTVRGTEKCYPEKVVSGK---DKPVVRSLS------KRP 2112
Db   1166 DLI--RMTYRSPIDVAHYLN----------YKSFYTGGRITTKNEVVRSMSVVGPRNVRP 1213

Qy   2113 IGTTDDVAEINADVYLCMTQLEKSDMKRSLKGKGKETPVMTVHEAQGKTFSDVVLFRTKK 2172
Db   1214 MTSVYSVPYVPGVQYLTFTQTEKDDLFKALRSKG-HVNVNTVHETQGKTFDDVILVRLKT 1272

Qy   2173 ADDSLF----TKQPHILVGLSRHTRSLV-YAALSSKL 2204
Db   1273 TENEIYPGGRNSKPYTIVGLTRHRRSLVYYTAIEDRL 1309
```

RESULT 15
Q9WIE3
ID Q9WIE3 PRELIMINARY; PRT; 1350 AA.
AC Q9WIE3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE REPLICASE.
OS Chinese wheat mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Furovirus.
OX NCBI_TaxID=83544;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YANTAI, CHINA;
RX MEDLINE=99281891; PubMed=10355760;
RA Diao A., Chen J., ye R., Zheng T., Yu S., Antoniw J.F., Adams M.J.;
RT "Complete sequence and genome properties of Chinese wheat mosaic
RT virus, a new furovirus from China.";
RL J. Gen. Virol. 80:1141-1145(1999).
DR EMBL; AJ012005; CAB41770.1; -.
DR INTERPRO; IPR000606; -.
DR INTERPRO; IPR002588; -.
DR PFAM; PF01443; Viral_helicase1; 1.
DR PFAM; PF01660; Vmethyltransf; 1.
SQ SEQUENCE 1350 AA; 153005 MW; F8D84D51BD426B1F CRC64;

Query Match 4.0%; Score 466; DB 12; Length 1350;
Best Local Similarity 19.1%; Pred. No. 3e-23;
Matches 372; Conservative 240; Mismatches 614; Indels 722; Gaps 79;

```
Qy    398 ASVANALNQEEVYSSVVSSITNRLVLRDQSALLSHLDTKLCDMFSQR--DAMIREKPSHR 455
Db      6 SSILGIISEEDVIRAAISTSATKFG--------SQLHSTVCDHVKETYLDAMERQKTKKK 57

Qy    456 CDV--FLKPRE-REKLRELFPELSIQFSDSVRSSHPFANAMRSCFNGIFSRRCGNVCF-FD 512
Db     58 IDVRRDLSQEQLQLLNDLYPERHIVSSNCERGTHSFAAASRKIETDLLLSRIPKKSWVYD 117

Qy    513 IGGSFTYHVKAGH-VNCHVCNPVLDVKDVKRRINEILFLSTAGGDSYV--SSDLLTEAAS 569
Db    118 IGGNWATHIKRNDGRRVHCCCPTIDIRDSARKT-----VSWASIEKYLDEKDEIPTEVGE 172

Qy    570 --KSVS----------YCSRESQNCDSRADAGFMV---DVYDISPQ 600
```



Search completed: January 31, 2001, 15:35:03
Job time: 548 sec

This Page Blank (uspto)